U.S. Application No. Unknown | International Application No. PCT/US98/27364 | Attorney Docket No. NIH220.001APC

Date: Herewith

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 USC 371

International Application No.:	PCT/US98/27364
International Filing Date:	December 23, 1998
Priority Date Claimed:	December 23, 1997 (Appln. No. 60/068,655)
Title of Invention:	IMMUNIZATION FOR EBOLA VIRUS INFECTION
Applicant(s) for DO/EO/US:	Gary J. Nabel and Anthony Sanchez

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. (X) This is a **FIRST** submission of items concerning a filing under 35 USC 371.

2. (X) This express request to begin national examination procedures (35 USC 371(f)) at any time rather than delay examination until the expiration of the applicable time limit set in 35 USC 371(b) and PCT Articles 22 and 39(1).

3. (X) A proper Demand for International Preliminary Examination was made by the 19th month from the earliest claimed priority date.

4. (X) A copy of the International Application as filed (35 USC 371(c)(2))

   a) () is transmitted herewith (required only if not transmitted by the International Bureau).
   b) () has been transmitted by the International Bureau.
   c) () A copy of Form PCT/1B/308 is enclosed.
   d) (X) is not required, as the application was filed in the United States Receiving Office (RO/US).

5. (X) Amendments to the claims of the International Application under PCT Article 19 (35 USC 371(c)(3))

   a) () are transmitted herewith (required only if not transmitted by the International Bureau).
   b) () have been transmitted by the International Bureau.
   c) () have not been made; however, the time limit for making such amendments has NOT expired.
   d) (X) have not been made and will not be made.

6. (X) An oath or declaration of the inventor(s) (35 USC 371(c)(4)).

7. (X) A copy of the International Preliminary Examination Report with any annexes thereto, such as any amendments made under PCT Article 34.

8. (X) A translation of the annexes, such as any amendments made under PCT Article 34, to the International Preliminary Examination Report under PCT Article 36 (35 USC 371(c)(5)).

9. (X) International Application as published.

10. (X) Copy of International Search Report and copies of the references cited therein.

11. (X) Petition for Revival of Patent Application.

518 Rec'd PCT/PTO 17 AUG 2001

U.S. Application No.
Unknown

International Application No.
PCT/US98/27364

Attorney Docket No.
NIH220.001A PC

09/913909

Date: Herewith

12. (X) A return prepaid postcard.

13. (X) The following fees are submitted:

				FEES
	BASIC FEE			$100
CLAIMS	**NUMBER FILED**	**NUMBER EXTRA**	**RATE**	
Total Claims	37 - 20 =	17 ×	$18	$306
Independent Claims	5 - 3 =	2 ×	$80	$160
Multiple dependent claims(s) (if applicable)			$270	$270
	TOTAL OF ABOVE CALCULATIONS			$836
Reduction by 1/2 for filing by small entity (if applicable). Verified Small Entity statement must also be filed. (NOTE 37 CFR 1.9, 1.27, 1.28)				$0
	TOTAL NATIONAL FEE			$836
	TOTAL FEES ENCLOSED			$2206

14. (X) A check in the amount of $2206 to cover the above fees is enclosed, $1240 petition fee, and $130 late filing of oath/declaration fee.

15. (X) The Commissioner is hereby authorized to charge only those additional fees which may be required, now or in the future, to avoid abandonment of the application, or credit any overpayment to Deposit Account No. 11-1410.

NOTE: Since an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) is filed herewith to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Nancy W. Vensko
Reg. No. 36,298
Customer No. 20,995

O:\DOCS\NWV\NWV-4769.DOC
070301

IMMUNIZATION FOR EBOLA VIRUS INFECTION

FIELD OF THE INVENTION

The present invention relates generally to viral vaccines and, more particularly, to Ebola virus vaccines and methods of protecting against disease caused by infection with Ebola virus.

BACKGROUND OF THE INVENTION

The Ebola viruses, and the genetically-related Marburg virus, are filoviruses associated with outbreaks of highly lethal hemorrhagic fever in humans and primates in North America, Europe, and Africa. Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996); Peters, C.J. et al , *Semin. Virol.* 5:147-154 (1994). Ebola viruses are negative-stranded RNA viruses comprised of four subtypes, including those described in the Zaire, Sudan, Reston, and Ivory Coast episodes. Sanchez, A. et al., *PNAS (USA)* 93:3602-3607 (1996). Although several subtypes have been defined, the genetic organization of these viruses is similar, each containing seven linearly arrayed genes. Among the viral proteins, the envelope glycoprotein exists in two alternative forms, a 50-70 kilodalton (kDa) secreted protein of unknown function encoded by the viral genome and a 130 kDa transmembrane glycoprotein generated by RNA editing that mediates viral entry. Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996); Sanchez, A. et al., *PNAS (USA)* 93:3602-3607 (1996). Other structural gene products include the nucleoprotein (NP), matrix proteins VP24 and VP40, presumed nonstructural proteins VP30 and VP35, and the viral polymerase (reviewed in Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996)). Although spontaneous variation of its RNA sequence does occur in nature, there appears to be less nucleotide polymorphism within Ebola subtypes than among other RNA viruses (Sanchez, A. et al., *PNAS (USA)* 93:3602-3607 (1996)), suggesting that immunization may be useful in protecting against this disease. Previous attempts to elicit protective immune responses against Ebola virus using traditional active and passive immunization approaches have, however, not succeeded. Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996); Clegg, J.C.S.

et al., *New Generation Vaccines.* (eds., Levine, M.M., Woodrow, G.C., Kaper, J.B.& Cobon, G.S.) 749-765 (New York, NY, Marcel Dekker, Inc. 1997); Jahrling, P.B. et al., *Arch. Virol. Suppl.* 11:135-140 (1996).

It would thus be desirable to provide a vaccine to protect against disease caused by infection with Ebola virus. It would further be desirable to provide methods of making and using said vaccine.

SUMMARY OF THE INVENTION

Ebola virus vaccines comprising nucleic acid molecules encoding Ebola viral proteins are provided. In one embodiment, the nucleic acid molecule encodes the transmembrane form of the viral glycoprotein (GP). In another embodiment, the nucleic acid molecule encodes the secreted form of the viral glycoprotein (sGP). In yet another embodiment, the nucleic acid molecule encodes the viral nucleoprotein (NP).

The present invention also provides methods for immunizing a subject against disease caused by infection with Ebola virus comprising administering to the subject an immunoeffective amount of an Ebola virus vaccine. Administration can be by any of the routes normally used for gene therapy. In a preferred method, the Ebola virus vaccine is administered by intramuscular injection. The genetic immunization methods of the present invention provide protective immunity against disease caused by infection with Ebola virus.

Additional objects, advantages, and features of the present invention will become apparent from the following description and appended claims, taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The various advantages of the present invention will become apparent to one skilled in the art by reading the following specification and subjoined claims and by referencing the following drawings.

Figures 1A and 1B are photographs showing expression of Ebola virus gene products in eukaryotic plasmid expression vectors.

Figure 1A. Expression vectors encoding the indicated viral gene products under regulation of the CMV immediate-early region 1 enhancer and promoter were prepared and transfected into 293 cells as previously described. Manthorpe, M. et al. *Hum. Gene. Ther.* 4:419-431 (1993); Sambrook, J., Fritch, E.F., & Maniatis, T. Cold Spring Harbor, N.Y. Cold Spring Laboratory Harbor Press, 1994. Cell extracts were prepared and analyzed by Western blot analysis for NP (left) or GP (right) using

relevant rabbit antisera and a secondary antibody, horseradish peroxidase conjugated donkey anti-rabbit IgG of a dilution of 1:5,000. Incubation with primary antibody was for 30 minutes at room temperature, and for 30 minutes at room temperature with secondary antibody. Immunocomplexes were then detected by chemiluminescence using super signal substrate reagents (Pierce) according to manufacturer's instructions.

Figure 1B. Generation of antibody response in mice immunized with the indicated vectors and analyzed by Western blot for NP, GP, and sGP as shown. Antisera from mice were tested at a dilution of 1:500 (NP), 1:50 (GP), or 1:50 (sGP), respectively, and developed with a secondary antibody (sheep anti-mouse, 1:5,000, Amersham Life Science) and chemiluminescence as in Figure 1A. The control vector used for immunization represents the expression vector plasmid with no insert. Manthorpe, M. et al., *Hum. Gene. Ther.* 4:419-431 (1993).

Figures 2A-2D are graphs showing the immune responses to NP and GP after genetic immunization in mice.

Figure 2A. Splenic lymphocytes from vector or NP-plasmid immunized mice were isolated approximately 6 weeks after the initial immunization and sensitized *in vitro* for 5 days with 10 U/ml hIL-2. Renca-NP cells sensitized splenocytes from vector-immunized or pCMV-NP immunized mice were used to detect CTL activity at the indicated effector:target ratios on Renca or Renca-NP cells (left, middle) or with allogeneic effector cells with Renca-NP to show that they are susceptible to lysis (right). Allogeneic effector cells were generated by incubating cells derived from mice with a C57Bl/6 background ($5x10^6$/ml) with irradiated Balb/c spleen cells ($5x10^6$/ml) in the presence of IL-2 (20 U/ml) for five days. The chromium release CTL assay with Renca-NP cells was performed in triplicate as previously described. Ohno, T. et al., *Gene. Ther.* 4:361-366 (1997).

Figure 2B. Balb/C female mice were immunized with the sGP plasmid expression vector and analyzed for their ability to lyse the syngeneic Renca cell line stably expressing GP. Isolation of stable transfectants, confirmation of expression, and CTL assay were performed as described (see, Specific Example, II. Methods). Renca-GP or sGP sensitized splenocytes from pCMV-GP or pCMV-sGP immunized mice were used to determine the specific killing of ^{51}chromium labeled Renca-GP cells at the indicated E/T ratios.

Figure 2C. Mice immunized with GP were analyzed for their ability to lyse a syngeneic CT26 cell stably expressing GP or CT26 vector control transduced line at the indicated E/T ratios.

Figure 2D. Cellular proliferative response in the indicated immunized mice. T cells, enriched or depleted (see, Specific Example, II. Methods), were incubated at 10^5 cells/ml with sGP condition media (25%). Background was determined with cells incubated in media from control transfected 293 cells and subtracted from proliferation seen in sGP-containing supernatants.

Figures 3A-3C are graphs showing immunization with sGP or GP expression plasmids induces T cell responses to sGP in guinea pigs.

Figures 3A-3C. Cell-mediated immunity in guinea pigs was analyzed by performing assays to detect cell proliferation to control or GP antigen (A) or T-cell growth factor production in response to the indicated antigens. The culture supernatants containing these antigens were prepared as previously described (Bottomly, K. et al., Measurement of human and murine interleukin 2 and interleukin 4. in *Current Protocols in Immunology.* (eds., Coligan, J.E., Kruisbeek, A.M., Margulies, D.H., Shevach, E.M.& Strober, W.) 6.3.1-6.3.12 (New York, John Wiley & Sons, Inc. 1992); Arai, H. et al., *Nat. Med.* 3:843-848 (1997)), and included at a final concentration of 10% (volume/volume). In A, cell numbers refer to the concentration of spleen cells per ml in the ^{3}H-thymidine proliferation assay. In B, supernatants from A, harvested at the time of the peak proliferative response to sGP, were incubated with primary guinea pig T cells maintained in 200 U/ml of human IL-2. The percent maximal response refers to the magnitude of stimulation in response to the indicated stimuli relative to supernatants from 24 hour concanaval (in A-stimulated cells (2 μg/ml)). The requirement of T lymphocytes in guinea pig spleen cells for the proliferative response to sGP, performed as described in Specific Example, II. Methods, is shown (C).

Figures 4A-4F are photographs showing the immunohistochemical analysis of Ebola virus antigens in liver, lung, and spleen from representative protected (GP-animal 3) or infected (vector-animal 2) guinea pigs.

Figures 4A-4F. Magnification: liver, 40x; lung, 20x; spleen, 20x.

Figure 5 is a schematic of the plasmid pVR 1012-GP(IC) (Ivory Coast strain of GP, SEQ ID NO: 1).

Figure 6 is a schematic of the plasmid pVR 1012-GP(S) (Sudan strain of GP, see SEQ ID NO: 2).

Figure 7 is a schematic of the plasmid pVR 1012-GP(Z) (Zaire strain of GP, see SEQ ID NO: 3).

Figure 8 is a schematic of the plasmid pVR 1012-sGP(Z) (Zaire strain of sGP, see SEQ ID NO: 4).

Figure 9 is a schematic of the plasmid pVR 1012-NP.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Ebola virus vaccines are provided comprising a nucleic acid molecule encoding an Ebola viral protein operatively-linked to a control sequence in a pharmaceutically acceptable carrier. In one embodiment, the nucleic acid molecule encodes the transmembrane form of the viral glycoprotein (GP). In another embodiment, the nucleic acid molecule encodes the secreted form of the viral glycoprotein (sGP). In yet another embodiment, the nucleic acid molecule encodes the viral nucleoprotein (NP).

The present invention further includes vaccines comprising nucleic acid molecules encoding Ebola viral proteins other than GP, sGP, and NP, *e.g.*, other structural gene products which elicit protective immunity from disease caused by infection with Ebola virus. The nucleic acid molecules of the vaccines of the present invention encode structural gene products of any Ebola viral strain including the Zaire, Sudan, Ivory Coast and Reston strains. Nucleic acid molecules encoding structural gene products of the genetically-related Marburg virus strains may also be employed. Moreover, the nucleic acid molecules of the present invention may be modified, *e.g.*, the nucleic acid molecules set forth herein may be mutated, as long as the modified expressed protein elicits protective immunity from disease caused by infection with Ebola virus. For example, the nucleic acid molecule may be mutated so that the expressed protein is less toxic to cells. The present invention also includes vaccines comprising a combination of nucleic acid molecules. For example, and without limitation, nucleic acid molecules encoding GP, sGP and NP of the Zaire, Sudan and Ivory Coast Ebola strains may be combined in any combination, in one vaccine composition.

The present invention also provides methods for immunizing a subject against disease caused by infection with Ebola virus comprising administering to the subject an immunoeffective amount of an Ebola virus vaccine. Methods of making and using Ebola virus vaccines are also provided by the present invention including the preparation of pharmaceutical compositions.

As referred to herein, the term "encoding" is intended to mean that the subject nucleic acid may be transcribed in a cell, *e.g.*, when the subject nucleic acid is linked to appropriate control sequences such as a promoter in a suitable vector (*e.g.*, an expression vector) and the vector is introduced into a cell. The nucleic acid molecules of the present invention may be DNA molecules, cDNA molecules or RNA molecules, and are preferably cDNA molecules. The term "operatively-linked" as used herein refers to functional linkage between a nucleic acid expression control sequence (such as a promoter) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence. Expression control sequences are known to those skilled in the art (see, *e.g.*, Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990)). Vectors which contain both a promoter and a cloning site to which an inserted piece of nucleic acid is operatively-linked to the promoter, are well known in the art and are generally referred to herein as "expression vectors" or "expression vector plasmids". Preferably, these vectors are capable of transcribing nucleic acid *in vitro* and *in vivo*. A preferred vector is the cytomegalovirus (CMV) expression vector which directs high levels of gene expression in muscle.

Nucleic acid molecules which hybridize under stringent conditions to the nucleic acid molecules described herein are also within the scope of the present invention. As will be appreciated by those skilled in the art, multiple factors are considered in determining the stringency of hybridization including species of nucleic acid, length of nucleic acid probe, T_m (melting temperature), temperature of hybridization and washes, salt concentration in the hybridization and wash buffers, aqueous or formamide hybridization buffer, and length of time for hybridization and for washes. An example of stringent conditions are DNA-DNA hybridization with a probe greater than 200 nucleotides in 5 x SSC, at 65°C in aqueous solution or 42°C in formamide, followed by washing with 0.1 x SSC, at 65°C in aqueous solution. (Other experimental conditions for controlling stringency are described in Maniatis, T. et al., *Molecular Cloning: a Laboratory Manual*, Cold Springs Harbor Laboratory, Cold Springs, N.Y. (1982) at pages 387-389 and Sambrook, J. et al., *Molecular Cloning: a Laboratory Manual*, Second Edition, Volume 2, Cold Springs Harbor Laboratory, Cold Springs, N.Y., at pages 8.46-8.47 (1989)).

It will be appreciated that administration of the vaccines of the present invention can be by any of the routes normally used for gene therapy. In a preferred

method, administration is by intramuscular injection, however, other procedures for transfecting cells may also be employed, such as transfection using calcium phosphate coprecipitation, liposome-mediated transfection, plasmid and viral vector-mediated transfection and DNA protein complex-mediated transfection. Viral vector-mediated transfection includes, without limitation, the use of retroviral, replication-deficient retroviral, adenoviral and adeno-associated viral vectors. Cells transfected by the vaccines in the context of *ex vivo* gene therapy can also be administered.

It will be appreciated that more than one route of administering the vaccines of the present invention may be employed either simultaneously or sequentially (*e.g.*, boosting). In addition, the vaccines of the present invention may be employed in combination with traditional immunization approaches such as employing protein antigens, vaccinia virus and inactivated virus, as vaccines. Thus, in one embodiment, the vaccines of the present invention are administered to a subject (the subject is "primed" with a vaccine of the present invention) and then a traditional vaccine is administered (the subject is "boosted" with a traditional vaccine). In another embodiment, a traditional vaccine is first administered to the subject followed by administration of a vaccine of the present invention. In yet another embodiment, a traditional vaccine and a vaccine of the present invention are co-administered.

Immunogenicity may be significantly improved if the vaccines of the present invention are co-administered with an immunostimulatory agent or adjuvant. Adjuvants enhance immunogenicity but are not necessarily immunogenic themselves. Immunostimulatory agents or adjuvants have been used for many years to improve the host immune responses to, for example, vaccines. Adjuvants may thus be employed to enhance the immunogenicity of the vaccines of the present invention, as well as the immunogenicity of traditional vaccines. Suitable adjuvants are well known to those skilled in the art and include, without limitation, aluminum phosphate, aluminum hydroxide, QS21, Quil A, derivatives and components thereof, calcium phosphate, calcium hydroxide, zinc hydroxide, a glycolipid analog, an octodecyl ester of an amino acid, a muramyl dipeptide, polyphosphazene, a lipoprotein, ISCOM matrix, DC-Chol, DDA, and other adjuvants and bacterial toxins, components and derivatives thereof.

The vaccines of the present invention may also be co-administered with cytokines to further enhance immunogenicity. The cytokines may be administered by methods known to those skilled in the art, *e.g.*, as a nucleic acid molecule in plasmid form or as a protein or fusion protein.

Upon inoculation with a pharmaceutical composition as described herein, the immune system of the host responds to the vaccine by producing antibodies, both secretory and serum, specific for Ebola virus proteins. As a result of the vaccination, the host becomes at least partially or completely immune to Ebola virus infection, or resistant to developing moderate or severe disease caused by Ebola virus infection. Although Ebola virus infection and disease caused thereby are discussed in detail herein, it will be appreciated that the vaccines and methods of the present invention may be employed to immunize a subject against hemorrhagic fever generally, such as that caused by infection by the genetically-related Marburg virus.

Pharmaceutical compositions comprising the nucleic acid molecules encoding Ebola viral proteins described herein, either alone or in combination, and a pharmaceutically acceptable carrier, are also provided by the present invention. As used herein, the phrase "pharmaceutically acceptable carrier" encompasses any of the standard pharmaceutical carriers, such as those suitable for parenteral administration, such as, for example, by intramuscular, intraarticular (in the joints), intravenous, intradermal, intraperitoneal, and subcutaneous routes. Examples of such formulations include aqueous and non-aqueous, isotonic sterile injection solutions, which contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives.

Formulations suitable for oral administration can consist of (a) liquid solutions, such as an effective amount of the vaccine dissolved in diluents, such as water, saline or PEG 400; (b) capsules, sachets or tablets, each containing a predetermined amount of the vaccine, as liquids, solids, granules or gelatin; (c) suspensions in an appropriate liquid; (d) suitable emulsions; and (e) polysaccharide polymers such as chitians. The vaccine, alone or in combination with other suitable components, may also be made into aerosol formulations to be administered via inhalation, *e.g.*, to the bronchial passageways. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Suitable formulations for rectal administration include, for example, suppositories, which consist of the vaccine with a suppository base. Suitable suppository bases include natural or synthetic triglycerides or paraffin hydrocarbons. In addition, it is also possible to use gelatin rectal capsules which consist of a

combination of the vaccine with a base, including, for example, liquid triglycerides, polyethylene glycols, and paraffin hydrocarbons.

Each carrier must be "acceptable" in the sense of being compatible with the other ingredients of the formulation and not injurious to the recipient, *e.g.*, the patient. The formulations can be presented in unit-dose or multi-dose sealed containers, such as ampules or vials and may be prepared by any method known in the art.

Pharmaceutical compositions comprising any of the nucleic acid molecules encoding Ebola viral proteins of the present invention are useful to immunize a subject against disease caused by Ebola virus infection. Thus, this invention further provides methods of immunizing a subject against disease caused by Ebola virus infection, *e.g.*, hemorrhagic fever, comprising administering to the subject an immunoeffective amount of a pharmaceutical composition of the invention. This subject may be an animal, for example a mammal, such as a primate or preferably a human.

The vaccines of the present invention are also suitable for veterinary immunization. The vaccines of the present invention comprising nucleic acid molecules encoding Ebola virus structural gene products from the Reston strain, which is known to infect animals, are particularly useful in such veterinary immunization methods.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be therapeutically effective, immunogenic and protective. The quantity to be administered depends on the subject to be treated, including, for example, the capacity of the immune system of the individual to synthesize antibodies, and, if needed, to produce a cell-mediated immune response. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and may be monitored on a patient-by-patient basis. However, suitable dosage ranges are readily determinable by one skilled in the art and generally range from about 300 μg to about 4-5 mg. The dosage may also depend, without limitation, on the route of administration, the patient's state of health and weight, and the nature of the formulation.

Methods of immunizing a subject against multiple strains of Ebola virus are further provided herein. The nucleic acid molecules encoding Ebola viral proteins of the present invention may be combined with nucleic acid molecules encoding other viral proteins of other virus strains to achieve protection against multiple strains of

Ebola virus. Typically the vaccines will be in an admixture and administered simultaneously, but may also be administered separately.

In some instances it may be desirable to combine the Ebola virus vaccines of the present invention with vaccines which induce protective responses to other agents, particularly other viruses. For example, the vaccine compositions of the present invention can be administered simultaneously, separately or sequentially with other genetic immunization vaccines such as those for influenza (Ulmer, J.B. et al., *Science* 259:1745-1749 (1993); Raz, E. et al., *PNAS (USA)* 91:9519-9523 (1994)), malaria (Doolan, D.L. et al., *J. Exp. Med.* 183:1739-1746 (1996); Sedegah, M. et al., *PNAS (USA)* 91:9866-9870 (1994)), and tuberculosis (Tascon, R.C. et al., *Nat. Med.* 2:888-892 (1996)).

It will also be appreciated that single or multiple administrations of the vaccine compositions of the present invention may be carried out. For example, subjects who are particularly susceptible to Ebola virus infection may require multiple immunizations to establish and/or maintain protective immune responses. Levels of induced immunity can be monitored by measuring amounts of neutralizing secretory and serum antibodies, and dosages adjusted or vaccinations repeated as necessary to maintain desired levels of protection.

This invention also provides kits comprising the vaccines of the present invention. For example, kits comprising a vaccine and instructions for use are within the scope of this invention.

The vaccines and methods of the present invention evoke a protective immune response and do not lead to immunopotentiation or exacerbated disease. The vaccines lack transmissibility, are genetically stable and induce protective levels of humoral and cell-mediated immunity.

In order to more fully demonstrate the advantages arising from the present invention, the following example is set forth. It is to be understood that the following is by way of example only and is not intended as a limitation on the scope of the invention.

SPECIFIC EXAMPLE

I. RESULTS

Immune response to viral gene products in mice. To characterize immune responses to selected Ebola virus proteins, eukaryotic expression vector plasmids were injected into mice. The cytomegalovirus (CMV) immediate early region 1 enhancer was used to stimulate transcription because it directs high levels of gene

expression in muscle. Manthorpe, M. et al., *Hum. Gene. Ther.* 4:419-431 (1993). cDNAs encoding an abundant structural protein, the major viral nucleocapsid phosphoprotein (NP), the secreted glycoprotein (sGP), or the membrane-associated glycoprotein (GP) were inserted. Alternative forms of GP were chosen because it had been postulated that the transmembrane protein contained a protein sequence motif also found in oncogenic retroviruses that might suppress immune responses. Burkreyev, A.A. et al., *FEBS. Lett.* 323:183-187 (1993); Cianciolo, G.J. et al., *Science* 230:453-455 (1985); Harris, D.T. et al., *J. Immunol.* 138:889-894 (1987); Volchkov, V.E. et al., *FEBS. Lett.* 305:181-184 (1992); Sanchez, A. et al., *Virus. Res.* 29:215-240 (1993). Expression of the relevant proteins was confirmed after transfection of the human renal epithelial cell line, 293, by immunoblotting with antisera to these gene products (Fig. 1A). For NP, the expected full-length 104 kDa protein normally produced by the virus was seen, together with lower molecular weight species likely generated from truncated protein or degradation products described previously. Sanchez, A. et al., *Virology* 170:81-91 (1989). Similarly, expression of sGP and GP revealed a heterogeneous pattern whose sizes correlated with the expected products of cleavage or post-translational carbohydrate modification. Sanchez, A. et al., *PNAS (USA)* 93:3602-3607 (1996).

These plasmids were injected into mice to characterize their ability to induce humoral and cellular immune responses to the relevant viral proteins. Three injections, each with 50 μg of plasmid DNA in saline (100 μl), were performed at two-week intervals in Balb/C female mice (6-8 week old, Charles River). Serum from immunized recipients were examined for antibody responses. An antibody response to the viral NP gene product was readily detectable (Fig. 1B), with titers of $\geq$ 1:16,000 by Western blot analysis. The titer of antibody induced in response to injection with plasmids encoding the viral glycoproteins was lower. After immunization with GP, no antibody was detectable by Western blotting, while immunization with sGP induced an antibody response (Fig. 1B). The more sensitive ELISA (Ksiazek, T.G., *Lab. Anim.* 20:34-46 (1991); Ksiazek, T.G. et al., *J. Clin. Microbiol.* 30:947-950 (1992)) did allow detection of antibodies to both GP and sGP at titers of 1:400 and 1:1,200, respectively. Cytolytic T cell (CTL) responses to these viral proteins were analyzed next. Despite the substantial humoral immune response to NP, minimal CTL activity was detected against syngeneic cells expressing this viral protein (Fig. 2A). In contrast, genetic immunization with sGP, which elicited a weaker antibody response, induced a marked cytolytic T cell response to cells expressing GP (Fig. 2B).

Immunization with the GP plasmid also induced a significant CTL response to GP (Fig. 2C). These data suggested that both the secreted and transmembrane form of the protein could be processed for antigen presentation and the transmembrane form was a target for recognition by these cytolytic T cells. Finally, antigen-specific T cell proliferation to sGP was also observed in GP and sGP but not plasmid control injected mice (Fig. 2D).

The ability of antibodies detected in mouse sera after immunization to neutralize virus was tested in an *in vitro* infection assay. McCormick, J.B. et al., *J. Infect. Dis.* 147:264-267 (1983). In no case was neutralization of infectivity observed, even at titers of 1:10 (data not shown), despite the documented presence of antibody after NP and sGP immunization by Western blot analysis. Infectivity *in vitro* was thus not inhibited by Ebola-specific antibodies.

Immune function and viral challenge in guinea pigs. To determine whether the *in vivo* immune responses could protect against viral infection, virus was adapted to grow in guinea pigs. Though this species is not well-suited to analysis of immune function, infection in adult mice has not been successful. Moreover, infection in guinea pigs, used originally to propagate virus from infected humans, is a well-established animal model for the human disease. Infection gives rise to a syndrome of hemorrhagic fever with levels of virus, organ pathology, and infection of reticuloendothelial and mononuclear cells comparable to humans. Bowen, E.T.W. et al., *Lancet* 1:571-573 (1977).

Two groups of immunized guinea pigs were studied. Animals were injected intramuscularly with the relevant expression vector plasmids, and the response to infection in groups immunized with either sGP, GP, NP, or control plasmids was observed. In the first group, animals were challenged within 2 months after the initial immunization, at which time the antibody titers were high, ranging from 1:1,600 to >1:25,000 (Table 1A). In these animals, nearly complete protection from lethal challenge was observed in GP (6/6), sGP (5/6), and NP (4/4) immunized subjects, in contrast to controls (0/6). In a second group, guinea pigs were challenged four months after the initial immunization (Table 1B). As in the first group, all animals immunized with the control vector succumbed to infection within a week after virus challenge (n=4). In this group, antibody titers were lower, and three of the four guinea pigs immunized with NP succumbed to infection, with the single survivor appearing severely ill after 1 week, in contrast to the protective response with NP at the earlier time point after immunization in Group I. More effective protection was

seen in animals immunized with vector expressing GP, protection was noted in four of five animals challenged, with one survivor appearing weak but surviving the viral challenge. Similarly, three of the five animals immunized with sGP showed no ill effects following viral challenge. Protection in this group again correlated with the ability to sustain an effective immune response to GP or sGP. Together, all guinea pigs immunized with vectors expressing GP or sGP which had titers greater than 1:5,120 were resistant to infection (11/11) compared to 0/10 controls (p=0, by Fisher's exact test). Twelve of fourteen animals with antibody titers ≥2,560 survived viral challenge compared to controls (p=.00003, by Fisher's exact test). Similar to immunized mice, guinea pigs injected with GP or sGP plasmids were able to generate cell-mediated immune responses to the viral glycoprotein in addition to the antibody response. These responses were antigen-specific and T cell-dependent, as detected in sGP antigen-dependent spleen cell proliferation and T-cell growth factor assays (Fig. 3A-C). Thus, the ability to generate and sustain significant cellular immune responses *in vivo* correlated with protection from infection. Though antibody titer correlated with protection, cell-mediated immunity appeared necessary for protection since passive transfer of antibody to GP does not confer protection (Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996); Jahrling, P.B. et al., *Arch. Virol. Suppl.* 11:135-140 (1996)) and antisera from protected guinea pigs did not inhibit virus replication *in vivo* (n=3) or at a 1:10 dilution *in vitro* (data not shown). Since the Hartley guinea pig to which the virus has been adapted for growth is outbred, cellular adoptive transfer studies could not be performed.

TABLE 1 - Group I

Plasmid	Subject	ELISA(Pre)	ELISA(Post)	Viral Ag	Survival
GP	1	>1:25,600	1:12,800	-	Yes
GP	2	>1:25,600	1:25,600	-	Yes
GP	3	>1:25,600	1:25,600	-	Yes
GP	4	1:25,600	1:6,400	-	Yes
GP	5	1:25,600	1:12,800	-	Yes
GP	6	1:25,600	1:25,600	-	Yes
SGP	1	1:12,800	1:25,600	-	Yes
SGP	2	1:6,400	1:25,600	-	Yes
SGP	3	1:6,400	1:25,600	-	Yes

SGP	4	1:25,600	1:25,600	-	Yes
SGP	5	>1:25,600	1:12,800	-	Yes
SGP	6	1:1,600	Negative	+	No
NP	1	1:12,800	>1:25,600	-	Yes
NP	2	>1:25,600	1:25,600	-	Yes
NP	3	1:12,800	1:12,800	-	Yes
NP	4	1:25,600	1:25,600	-	Yes
Vector alone	1	Negative	Negative	+	No
Vector alone	2	Negative	n.d.	+	No
Vector alone	3	Negative	Negative	+	No
Vector alone	4	Negative	Negative	+	No
Vector alone	5	Negative	n.d.	+	No
Vector alone	6	Negative	n.d.	+	No

Guinea pigs were immunized on days 1, 14, 28, 42, and challenged on day 62.

TABLE 1 - Group II

Plasmid	Subject	ELISA(Pre)	ELISA(Post)	Viral Ag	Survival
GP	1	1:2,560	n.d.	+/f	No
GP	2	1:5,120	1:10,240	-	Yes
GP	3	1:10,240	1:10,240	-	Yes
GP	4	1:1,280	n.d.	+/f	No
GP	5	1:5,120	1:20,480	-	Yes (ill)
SGP	1	1:2,560	n.d.	+	No
SGP	2	1:10,240	1:5,120	+/f	Yes
SGP	3	1:10,240	1:81,920	-	Yes
SGP	4	1:2,560	1:5,120	-	Yes
SGP	5	1:640	n.d.	+	No
NP	1	n.d.	n.d.	+	No
NP	2	n.d.	n.d.	+	No
NP	3	n.d.	n.d.	+	No
NP	4	n.d.	Negative	+	Yes (ill)
Vector alone	1	Negative	n.d.	+	No
Vector alone	2	Negative	n.d.	+	No
Vector alone	3	Negative	n.d.	+	No
Vector alone	4	Negative	n.d.	+	No

Guinea pigs were immunized on days 1, 14, 42, and 112 and challenged on day 122.

n.d.=not done. Viral ag denotes presence of virus determined by immunohistochemistry (30) performed on spleen, liver, lung, kidney, and heart tissues; "+" = widespread systemic involvement of the mononuclear phagocyte system and to a lesser extent endothelial and parenchymal cells; "+/f" = focal involvement (seen in the spleen of SGP #2, the liver and spleen of GP #1, and the lung of GP#4) where rare sites of anti-Ebola antibody staining were detected.; "-" = no Ebola virus antigen detected in tissues.

ELISA determinations made prior to viral challenge (Pre) or at least 7 days after (Post) infection, respectively.

The surviving NP immunized animal (4) was found to have significant levels of virus in major organs by immunohistochemistry, and more than 5 logs of virus was detected in the serum and spleen, in contrast to GP and sGP animals where no virus was detected.

Histopathologic analysis of infection in immunized guinea pigs. Pathologic analysis revealed widespread tissue necrosis and dissemination of virus by immunohistochemistry, similar to human disease. Virus load correlated with susceptibility to infection with titers of $\geq 10^5$ in infected animals compared to undetectable levels in immunized survivors. In infected animals, the liver, lung, and spleen showed evidence of significant viral antigen by immunohistochemistry (Fig. 4, Table 1), and both reticuloendothelial and mononuclear phagotic involvement was observed.

Determination of antibody response in animals which survived virus challenge revealed increases in the immune response to viral proteins when initial titers were lower (Table 1). Less consistent increases in antibody titers were observed in the NP immunized animals. These data suggest that Ebola virus infection may stimulate immunity in survivors of a viral challenge when immune responses are not optimal.

II. METHODS

Plasmids. Plasmids containing the GP, sGP, or NP cDNAs (Sanchez, A. et al., *Virus. Res.* 29:215-240 (1993), Genbank) were used to subclone the relevant inserts into CMV expression vectors which utilized the bovine growth hormone polyadenylation sequence. Manthorpe, M. et al., *Hum. Gene. Ther.* 4:419-431 (1993). (see Figures 5-9 and SEQ ID NOS: 1-4). Briefly, for GP, plasmid pGEM-3Zf(-)-GP was digested with EcoR I, treated with the Klenow fragment of E. coli DNA polymerase, and digested with BamH I. The GP fragment was then inserted into the pCMV expression vector plasmid digested with BamH I, Klenow fragment and Bgl II. For sGP, the plasmid pCRII-sGP was digested with EcoR I, treated with Klenow

enzyme, and the resulting fragment inserted into the BamH I/Bgl II CMV plasmid which had been incubated with Klenow fragment, calf intestinal phosphatase (CIP), then phenol chloroform extract. For the NP expression vector, plasmid pSP64-NP2 (Sanchez, A. et al., *Virology* 170:81-91 (1989)) was digested with EcoR I, treated with Klenow enzyme, and digested with BamH I. The NP insert was cloned into CMV treated with BamH I, Klenow enzyme, followed by heat inactivation and Bgl II digestion.

Cell lines and transfectants. For stable transfectants, the relevant cDNAs were inserted into a CMV expression plasmid containing a neomycin resistant gene, pCMV-neo (H. Arai, unpublished data), which was digested with Xba I, and treated with CIP and Klenow enzyme. The EcoR I/ BamH I GP fragment from pGEM-3Zf(-)-GP, the EcoR I sGP fragment from pCRII-SGP, or the EcoR I/BamH I NP fragment from pSP64-NP2 was treated with Klenow enzyme and ligated to this plasmid backbone. These vectors were transfected into Renca or CT26 which was syngeneic to Balb/C mice using calcium phosphate and selected in 0.7 or 1mg/ml G418 for 2-6 weeks. Expression of GP, sGP, or NP from these vectors in Renca or CT26 cells was also confirmed by Western blot analysis (data not shown).

Cell proliferation assay. Spleen cells from male Hartley guinea pigs or Balb/C female mice (8-10 weeks) immunized with the indicated plasmid expression vectors were incubated with sGP or vector control supernatants (25% volume:volume) from transfected 293 cells at the indicated cell concentrations. T cell depletion was performed using the CT5 monoclonal antibody (Tan, B.T.G. et al., *Hybridoma* 4:115-124 (1985)) (Biosource, Camarillo, CA) for guinea pigs or anti-Thy 1.2 antibody in the mouse using immunomagnetic microbeads (Miltenyi Biotec, Inc., Auburn, CA).

Viral challenge in guinea pigs. Animals were immunized by injection of 100 μl (0.5 mg/ml) in each hind leg (two injections at each time point) with the indicated plasmid expression vectors. Animals were challenged by inoculation with a stock of Ebola virus (Zaire, 1976) that had been passaged once in vero E6 cells and serially passaged by intraperitoneal injection of spleen homogenates in Hartley guinea pigs seven times. Immunized guinea pigs were injected intraperitoneally with 0.5 ml of a 1:1,000 dilution of spleen cell homogenate in Hank's balanced salt solution 122 days after the initial plasmid DNA injection (1000 pfu). Survival was determined 10 days later at which times animals were sacrificed for serologic and pathologic analysis. ELISA, enzyme-linked immunosorbent assay (Volchkov, V.E. et al., *FEBS. Lett.* 305:181-184 (1992); Sanchez, A. et al., *Virus. Res.* 29:215-240 (1993)) on infected

cell supernatants and enriched viral extracts containing GP, sGP, or NP were performed as previously described.

III. DISCUSSION

Following the initial report that injection of plasmid DNA into muscle could direct the synthesis of recombinant proteins (Wolff, J.A. et al., *Science* 247:1465-1468 (1990)), the suggestion was made that this gene transfer approach may be useful for vaccination and was termed genetic immunization. Tang, D.C. et al., *Nature* 356:152-154 (1992). This approach has been applied to different infectious diseases, including influenza (Ulmer, J.B. et al., *Science* 259:1745-1749 (1993); Raz, E. et al., *PNAS (USA)* 91:9519-9523 (1994)), malaria (Doolan, D.L. et al., *J. Exp. Med.* 183:1739-1746 (1996); Sedegah, M. et al., *PNAS (USA)* 91:9866-9870 (1994)), and tuberculosis (Tascon, R.C. et al., *Nat. Med.* 2:888-892 (1996)) and has also been used to modulate antibody and cell-mediated immune responses in autoimmune and allergic diseases. Raz, E. et al., *PNAS (USA)* 90:4523-4527 (1993); Waisman, A. et al., *Nat. Med.* 2:899-905 (1996); McCormick, J.B. et al., *J. Infect. Dis.* 147:264-267 (1983); Border, W.A. et al., *Nat. Med.* 1:1000-1001 (1995).

The immune response to selected Ebola virus proteins after genetic immunization in mice was analyzed and their ability to protect against lethal infection in a susceptible animal model, the guinea pig, was tested. The immune analyses performed in different species suggest similar patterns of response, though the specific peptides which may be recognized by the immune system to confer protection in the guinea pig could differ from the mouse. Because the principles of MHC antigen presentation and recognition apply broadly across species (Monaco, J.J., *Immunol. Today* 13:173-179 (1992); Jorgensen, J.L. et al., *Annu. Rev. Immunol.* 10:835-873 (1992); Zinkernagel, R.M. et al., *Immunol. Today* 18:14-17 (1997)), the finding that protection was observed in different members of an outbred strain and that similar immune responses were seen in different species is not unexpected and suggests that this approach may be applicable to humans.

Immunization with plasmids encoding distinct viral proteins induced different antibody and cytolytic T cell responses. The broadest immune response was conferred by GP and sGP, which induced both cellular and humoral immunity to the membrane-associated GP. In guinea pigs challenged with doses of virus that are otherwise lethal, sGP provided nearly equivalent protection to GP, with no significant difference between these groups. The ability of vectors expressing GP to confer immunity may be explained by the generation of lower molecular weight degradation

products (Fig. 1B) which could provide sufficient protein for antigen presentation to induce detectable, cellular, and humoral immune responses in guinea pigs.

Despite the fact that plasmid DNA injection has been shown to affect the immune response to different antigens in infectious and autoimmune diseases, the ability of individual gene products to protect against disease *in vivo* is not readily predictable. In particular, the rapid rates of Ebola virus replication and the poor immunogenicity of its proteins had previously rendered it resistant to immune interventions. Several attempts to confer protection with passive transfer of immunoglobulin were unsuccessful (Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996); Jahrling, P.B. et al., *Arch. Virol. Suppl.* 11:135-140 (1996)), in agreement with the finding set forth herein that antisera from protected animals fails to neutralize virus replication *in vitro*. Previous studies using formalin-fixed virus or purified viral proteins for immunization have also not proven effective. Peters, C.J. et al., *Filoviridae*: Marburg and Ebola Viruses. in *Fields Virology.* (eds., Fields, B.N., Knipe, D.M.& Howley, P.M.) 1161-1176 (Philadelphia, Lippincott-Raven, 1996); Clegg, J.C.S. & Sanchez, A. Vaccines against arenaviruses and filoviruses. in *New Generation Vaccines.* (eds., Levine, M.M., Woodrow, G.C., Kaper, J.B.& Cobon, G.S.) 749-765 (New York, NY, Marcel Dekker, Inc. 1997).

It is likely that traditional immunization approaches using protein antigens, vaccinia virus, or inactivated virus do not allow for appropriate uptake and presentation of viral antigens by dendritic or other antigen-presenting cells to induce protective immune responses. It has been shown recently that genetic immunization leads to production of recombinant protein(s) in muscle which are delivered to bone marrow-derived antigen-presenting cells. Iwasaki, A. et al., *J. Immunol.* 159:11-14 (1997); Doe, B. et al., *PNAS (USA)* 93:8578-8583 (1996); Corr, M. et al., *J. Exp. Med.* 184:1555-1560 (1996). Synthesis of Ebola glycoprotein after gene transfer apparently allows more efficient processing and presentation and the generation of immune responses not seen with virus or with viral vectors. GP is a large molecule which contains both T and B cell epitopes. Although antibody levels provide a surrogate marker of protection, the fact that passive transfer of antibody did not confer protection implies that immunoglobulin switching and synthesis is reflective of the T helper response to GP. Genetic immunization stimulates T helper cells to generate both CTL and B cell antibody responses to the virus. Although antibody production confirms effective immunization, a productive T cell response, likely involving T_H1 cell

stimulation, as shown by the T cell proliferation and CTL assays (Fig. 3), is needed for effective immunity. Taken together, these studies suggest that transcription and translation of viral genes in host cells by genetic immunization induces alternative, more effective, processing and antigen presentation which better stimulates immunity to Ebola virus. Since there are yet no effective antiviral agents, the ability to generate protective immunity by vaccination may prove useful in selected high risk populations, particularly in regions of ongoing outbreaks, and among medical and laboratory personnel exposed to the virus. Although it remains important to identify agents which treat acute infection, genetic immunization may help to limit the spread of this highly lethal infectious disease.

The foregoing discussion discloses and describes merely exemplary embodiments of the present invention. One skilled in the art will readily recognize from such discussion, and from the accompanying drawings and claims, that various changes, modifications and variations can be made therein without departing from the spirit and scope of the invention as defined in the following claims.

All references cited herein are incorporated by reference as if fully set forth.

WE CLAIM:

1. A pharmaceutical composition comprising a nucleic acid molecule encoding an Ebola virus structural gene product operatively-liked to a control sequence, in a pharmaceutically acceptable carrier.

2. The pharmaceutical composition of Claim 1, wherein the Ebola virus structural gene product is selected from the group consisting of the transmembrane form of virus glycoprotein, the secreted form of virus glycoprotein, virus nucleoprotein and combinations thereof.

3. The pharmaceutical composition of Claim 1, wherein the control sequence is a promoter.

4. The pharmaceutical composition of Claim 3, wherein the promoter is the CMV immediate-early region 1 promoter.

5. The pharmaceutical composition of Claim 1, further comprising an adjuvant.

6. The pharmaceutical composition of Claim 2, wherein the structural gene product is the transmembrane form of virus glycoprotein.

7. The pharmaceutical composition of Claim 2, wherein the structural gene product is the secreted form of virus glycoprotein.

8. The pharmaceutical composition of Claim 2, wherein the structural gene product is virus nucleoprotein.

9. A method of producing a vaccine against disease caused by infection by Ebola virus, comprising the steps of:

a) administering the pharmaceutical composition of Claim 1 to a test host to determine an amount and a frequency of administration thereof to elicit a protective immune response in said host; and

b) formulating said pharmaceutical composition in a form suitable for administration to a treatable host in accordance with said determined amount and frequency of administration.

10. A vaccine comprising a nucleic acid molecule encoding the transmembrane form of the Ebola virus glycoprotein operatively-linked to a control sequence, in a pharmaceutically acceptable carrier.

11. The vaccine of Claim 10, wherein the control sequence is a promoter.

12. The vaccine of Claim 11, wherein the promoter is the CMV immediate-early region 1 promoter.

13. The vaccine of Claim 10, further comprising an adjuvant.

14. A vaccine comprising a nucleic acid molecule encoding the secreted form of the Ebola virus glycoprotein operatively-linked to a control sequence, in a pharmaceutically acceptable carrier.

15. The vaccine of Claim 14, wherein the control sequence is a promoter.

16. The vaccine of Claim 15, wherein the promoter is the CMV immediate-early region 1 promoter.

17. The vaccine of Claim 14, further comprising an adjuvant.

18. A vaccine comprising a nucleic acid molecule encoding the Ebola virus nucleoprotein operatively-linked to a control sequence, in a pharmaceutically acceptable carrier.

19. The vaccine of Claim 18, wherein the control sequence is a promoter.

20. The vaccine of Claim 19, wherein the promoter is the CMV immediate-early region 1 promoter.

21. The vaccine of Claim 18, further comprising an adjuvant.

22. A method of immunizing a subject against hemorrhagic fever comprising the step of administering to the host an immunoeffective amount of the vaccine of any of Claims 10 to 21.

23. The method of Claim 22, wherein the hemorrhagic fever is caused by infection with Ebola virus.

24. The method of Claim 22, wherein the hemorrhagic fever is caused by infection with Marburg virus.

25. The method of Claim 22, wherein the host is a human and administration is by intramuscular injection.

26. The method of Claim 22, wherein the subject receives a second administration of an immunoeffective amount of a vaccine against disease caused by infection by Ebola virus or Marburg virus.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K 39/12, 45/00, 39/145, 39/155, 39/205 — A1

(11) International Publication Number: WO 99/32147

(43) International Publication Date: 1 July 1999 (01.07.99)

(21) International Application Number: PCT/US98/27364

(22) International Filing Date: 23 December 1998 (23.12.98)

(30) Priority Data:
60/068,655 — 23 December 1997 (23.12.97) — US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US — 60/068,655 (CON)
Filed on — 23 December 1997 (23.12.97)

(71) Applicant *(for all designated States except US)*: THE REGENTS OF THE UNIVERSITY OF MICHIGAN [US/US]; Technology Management Office, Wolverine Tower, Room 2071, 3003 South State Street, Ann Arbor, MI 48109-1280 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: NABEL, Gary, J. [US/US]; 385 Meadow Creek Drive, Ann Arbor, MI 48105 (US). SANCHEZ, Anthony [US/US]; 1303 Summit Pte. Way, Atlanta, GA 30329 (US).

(74) Agents: SMITH, DeAnn, F. et al.; Harness, Dickey & Pierce, P.L.C., P.O. Box 828, Bloomfield Hills, MI 48303 (US).

(81) Designated States: CA, JP, US, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: IMMUNIZATION FOR EBOLA VIRUS INFECTION

(57) Abstract

Ebola virus vaccines comprising nucleic acid molecules encoding Ebola viral proteins are provided. In one embodiment, the nucleic acid molecule encodes the transmembrane form of the viral glycoprotein (GP). In another embodiment, the nucleic acid molecule encodes the secreted form of the viral glycoprotein (sGP). In yet another embodiment, the nucleic acid molecule encodes the viral nucleoprotein (NP). Methods for immunizing a subject against disease caused by infection with Ebola virus are also provided.

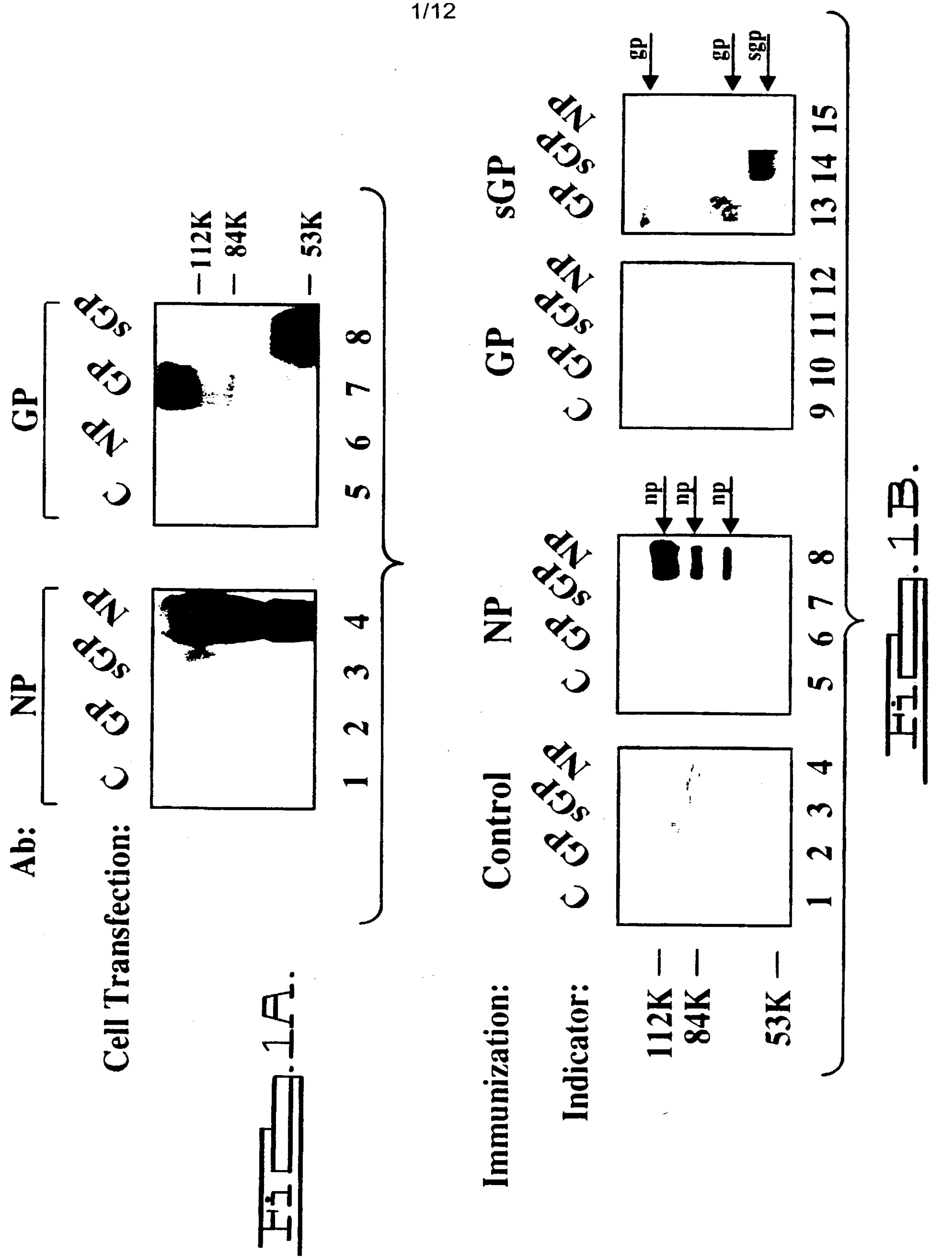

Ab:
NP
GP
Cell Transfection:
C GP sGP NP
C NP GP sGP
112K
84K
53K
1 2 3 4
5 6 7 8
Immunization:
Control
NP
GP
sGP
Indicator:
C GP sGP NP
C GP sGP NP
C GP sGP NP
C GP sGP NP
112K
84K
53K
1 2 3 4
5 6 7 8
9 10 11 12
13 14 15
np
np
np
gp
gp
sgp

Fig. 1A.

Fig. 1B.

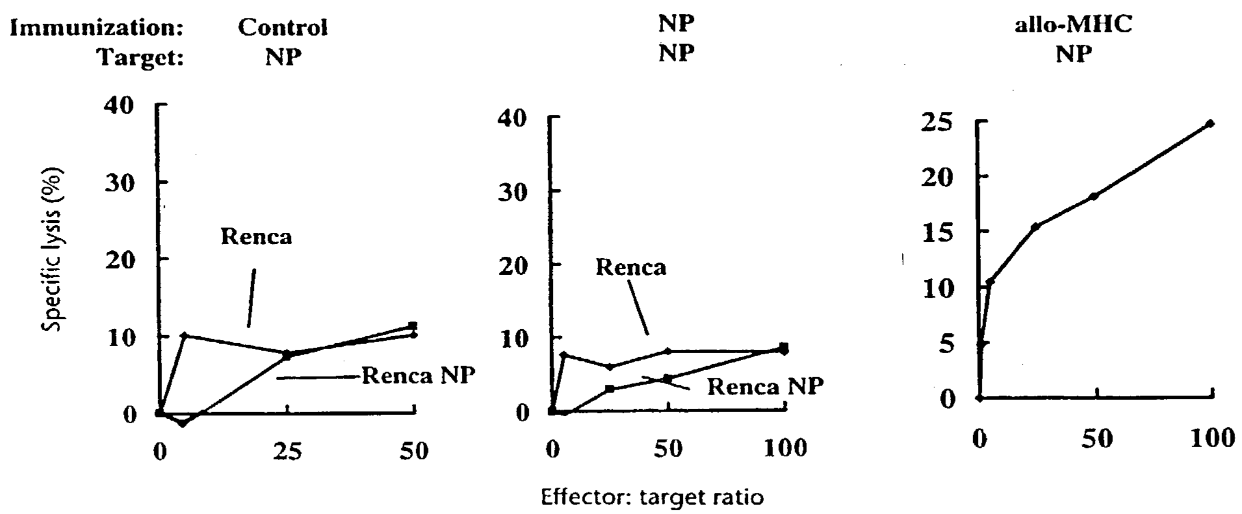


FIGURE 2A

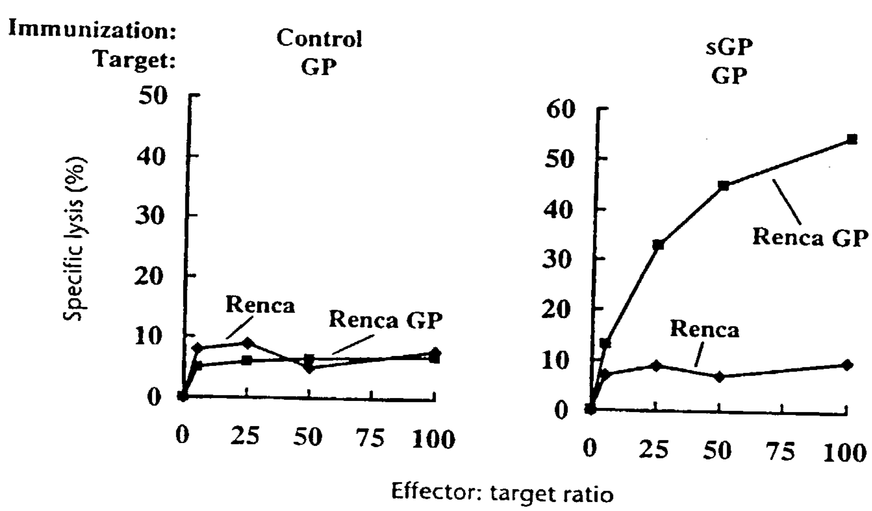


FIGURE 2B

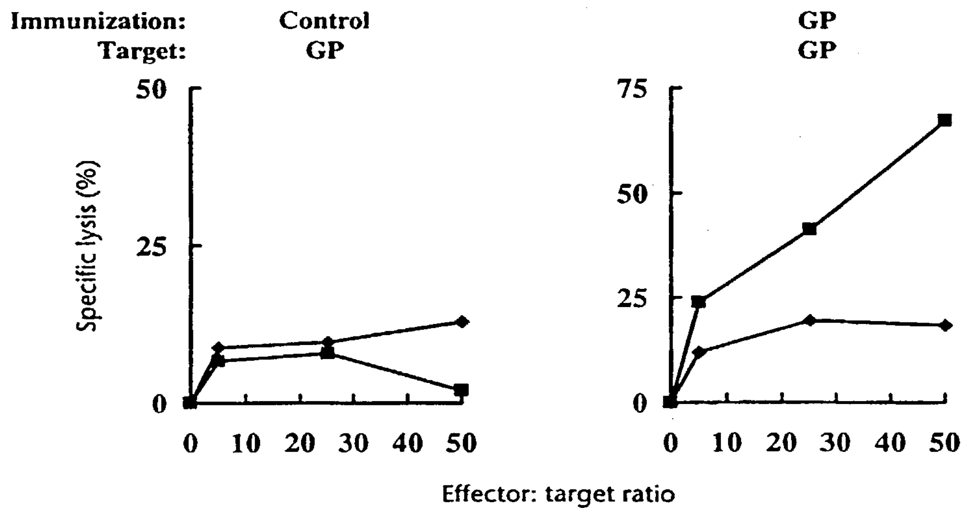


FIGURE 2C

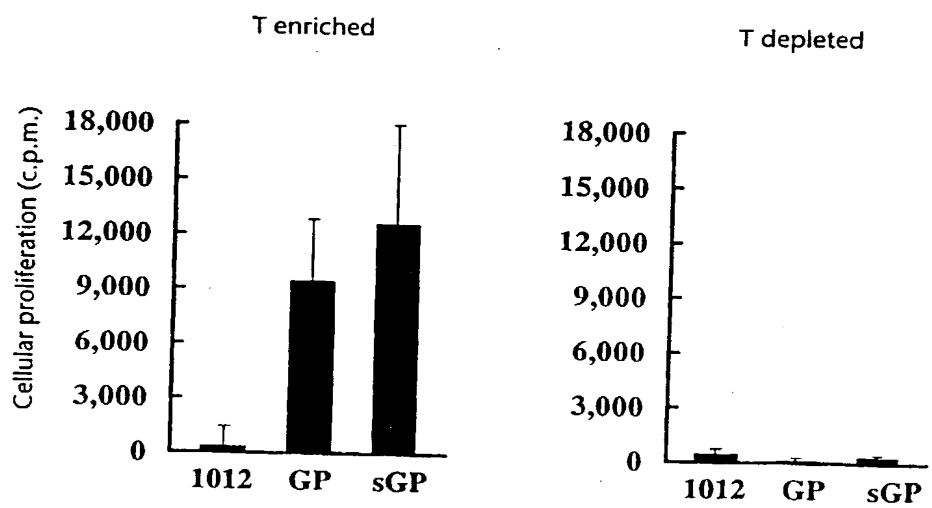


FIGURE 2D

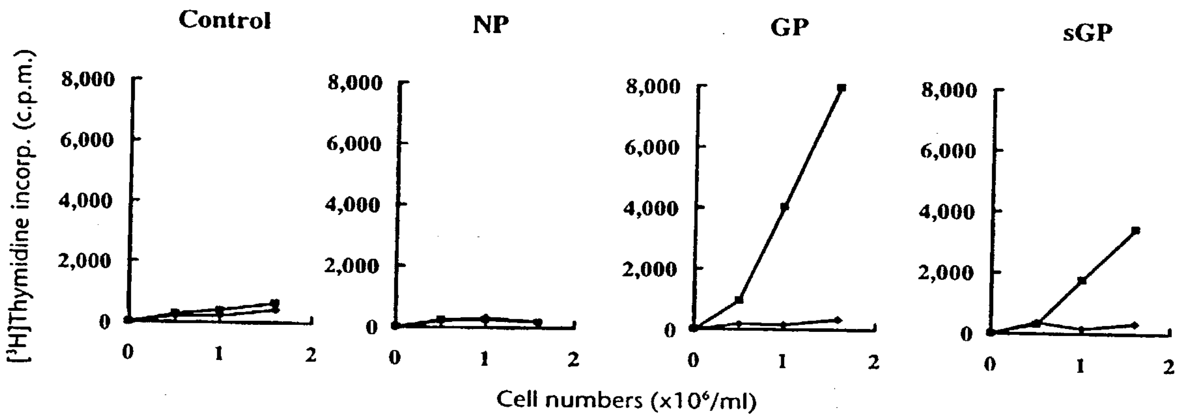


FIGURE 3A

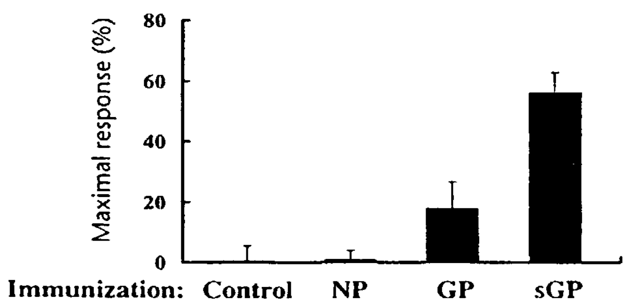


FIGURE 3B

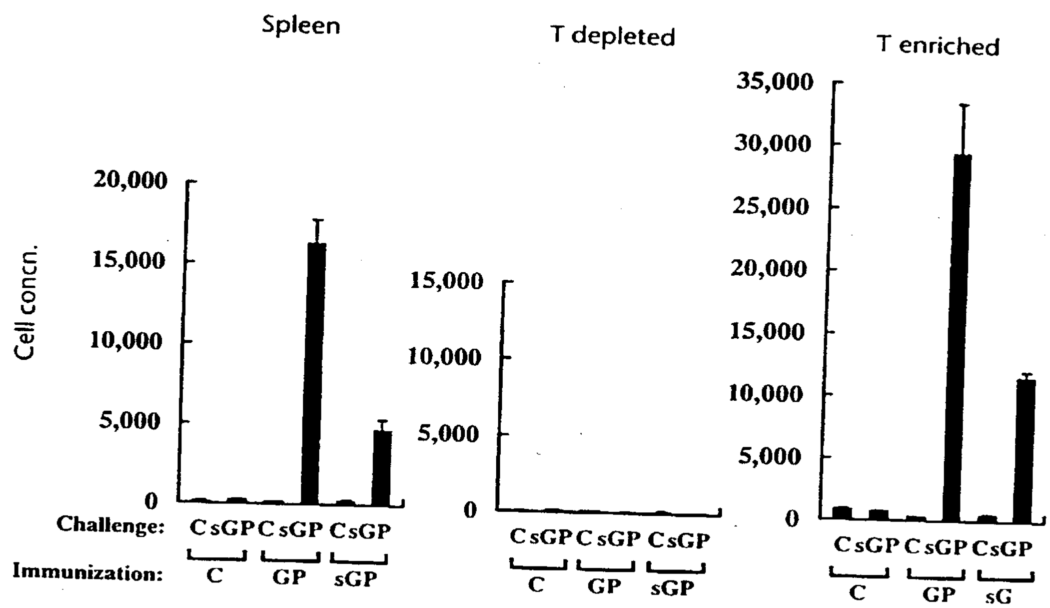

Spleen
T depleted
T enriched
Cell concn.
20,000
15,000
10,000
5,000
0
35,000
30,000
25,000
Challenge:
C sGP C sGP C sGP
Immunization:
C
GP
sGP
sG

FIGURE 3C

Protected

Liver:

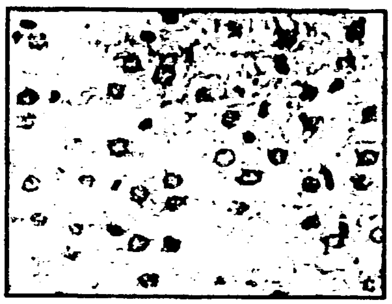


FIG. 4A.

Lung:

FIG. 4C.

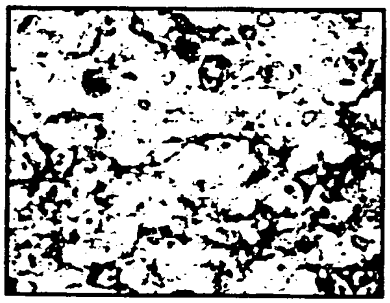
Spleen:

FIG. 4E.

Infected

FIG. 4B.

Liver:



Lung:

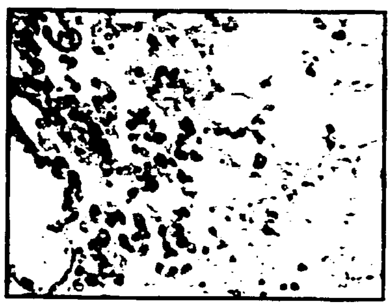


FIG. 4D.

Spleen:

FIG. 4F.

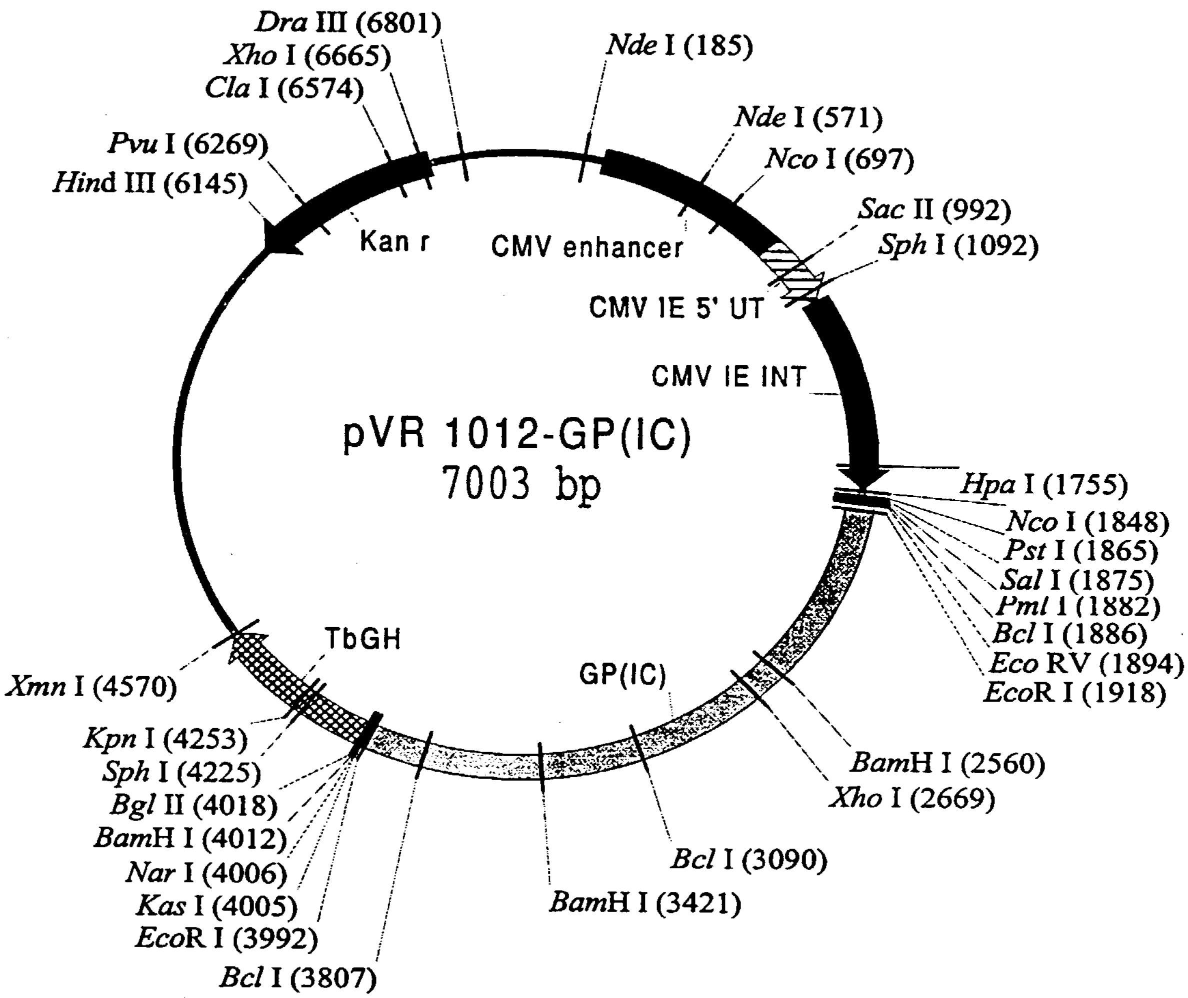

pVR 1012-GP(IC)
7003 bp
Dra III (6801)
Xho I (6665)
Cla I (6574)
Pvu I (6269)
Hind III (6145)
Kan r
Nde I (185)
Nde I (571)
Nco I (697)
CMV enhancer
Sac II (992)
Sph I (1092)
CMV IE 5' UT
CMV IE INT
Hpa I (1755)
Nco I (1848)
Pst I (1865)
Sal I (1875)
Pml I (1882)
Bcl I (1886)
Eco RV (1894)
EcoR I (1918)
BamH I (2560)
Xho I (2669)
GP(IC)
Bcl I (3090)
BamH I (3421)
TbGH
Xmn I (4570)
Kpn I (4253)
Sph I (4225)
Bgl II (4018)
BamH I (4012)
Nar I (4006)
Kas I (4005)
EcoR I (3992)
Bcl I (3807)

FIGURE 5

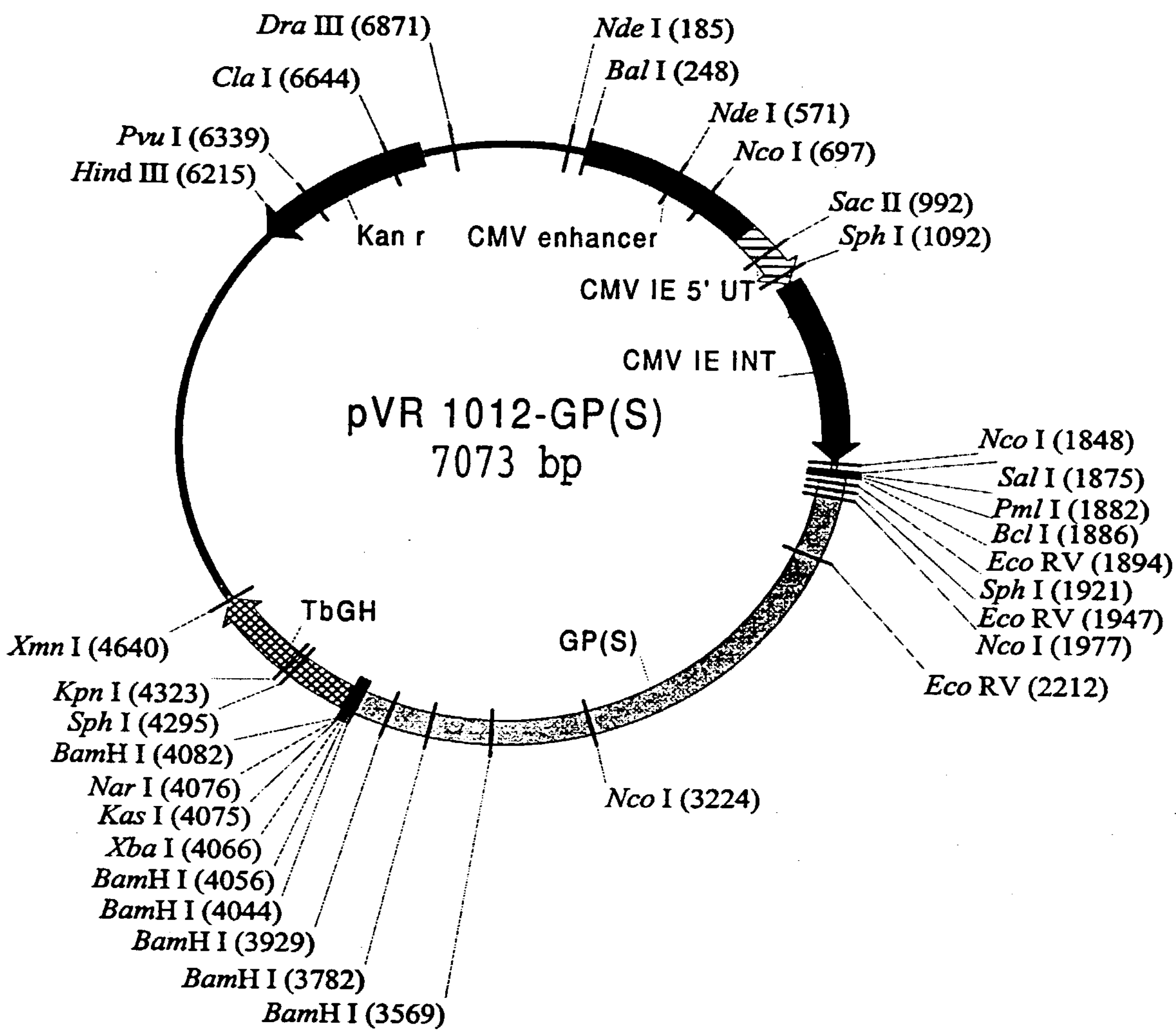

Dra III (6871)
Nde I (185)
Cla I (6644)
Bal I (248)
Nde I (571)
Pvu I (6339)
Nco I (697)
Hind III (6215)
Sac II (992)
Sph I (1092)
Kan r
CMV enhancer
CMV IE 5' UT
CMV IE INT
pVR 1012-GP(S)
7073 bp
Nco I (1848)
Sal I (1875)
Pml I (1882)
Bcl I (1886)
Eco RV (1894)
Sph I (1921)
Eco RV (1947)
Nco I (1977)
Eco RV (2212)
TbGH
GP(S)
Xmn I (4640)
Kpn I (4323)
Sph I (4295)
BamH I (4082)
Nar I (4076)
Kas I (4075)
Xba I (4066)
BamH I (4056)
BamH I (4044)
BamH I (3929)
BamH I (3782)
BamH I (3569)
Nco I (3224)

FIGURE 6

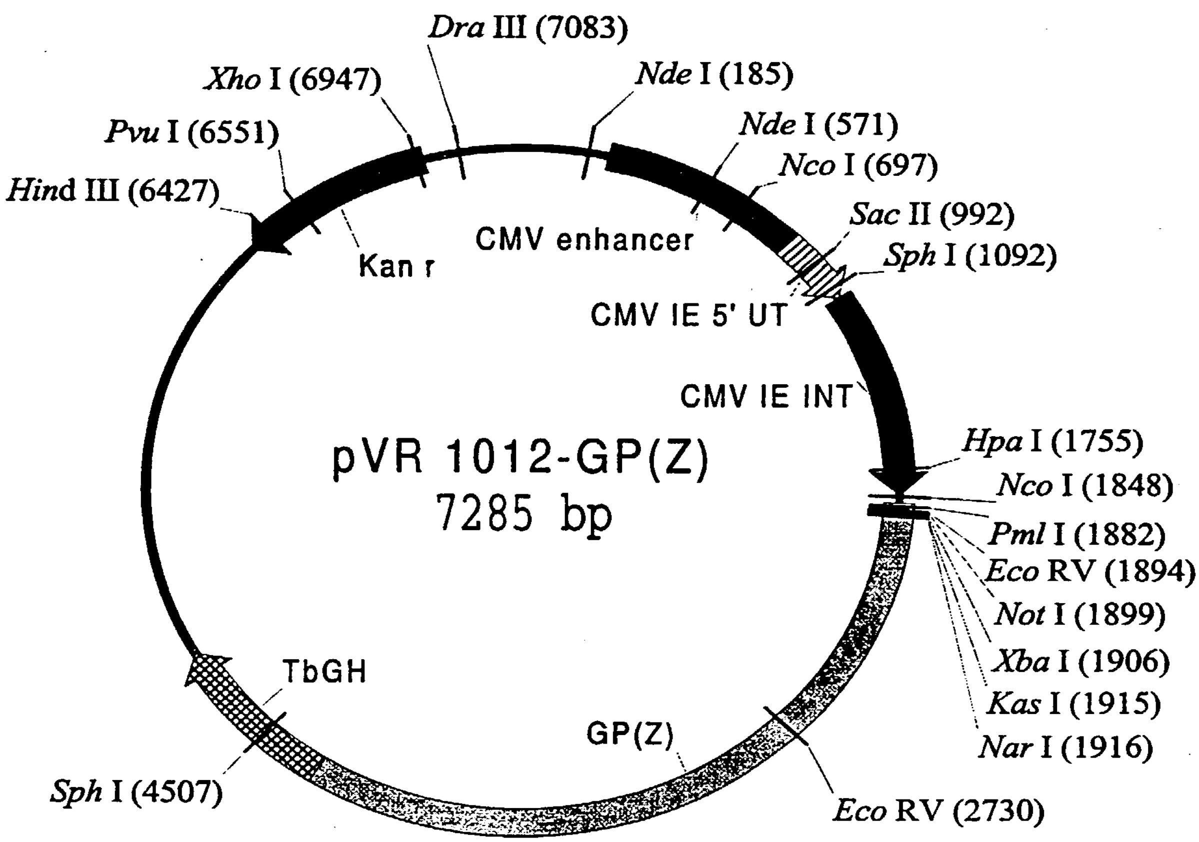

Dra III (7083)
Nde I (185)
Xho I (6947)
Nde I (571)
Pvu I (6551)
Nco I (697)
Hind III (6427)
Sac II (992)
CMV enhancer
Sph I (1092)
Kan r
CMV IE 5' UT
CMV IE INT
Hpa I (1755)
pVR 1012-GP(Z)
7285 bp
Nco I (1848)
Pml I (1882)
Eco RV (1894)
Not I (1899)
Xba I (1906)
TbGH
Kas I (1915)
GP(Z)
Nar I (1916)
Sph I (4507)
Eco RV (2730)

FIGURE 7

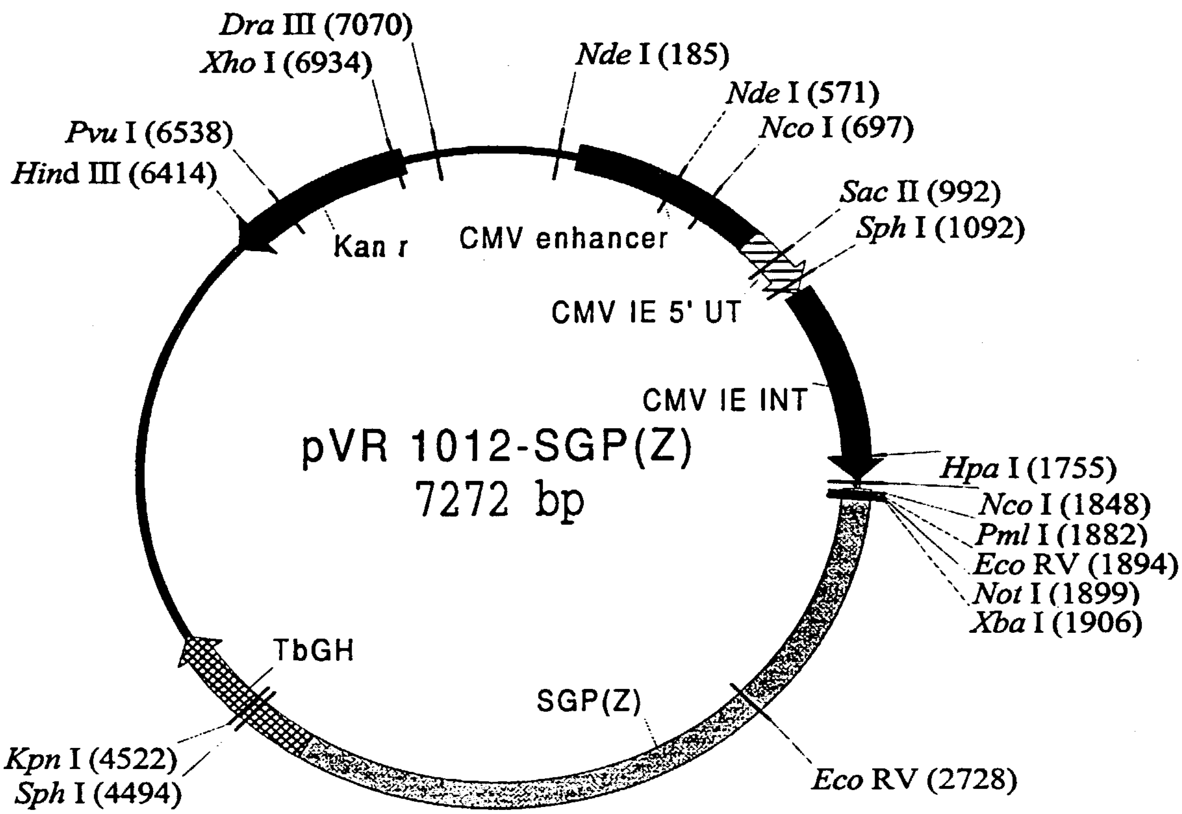
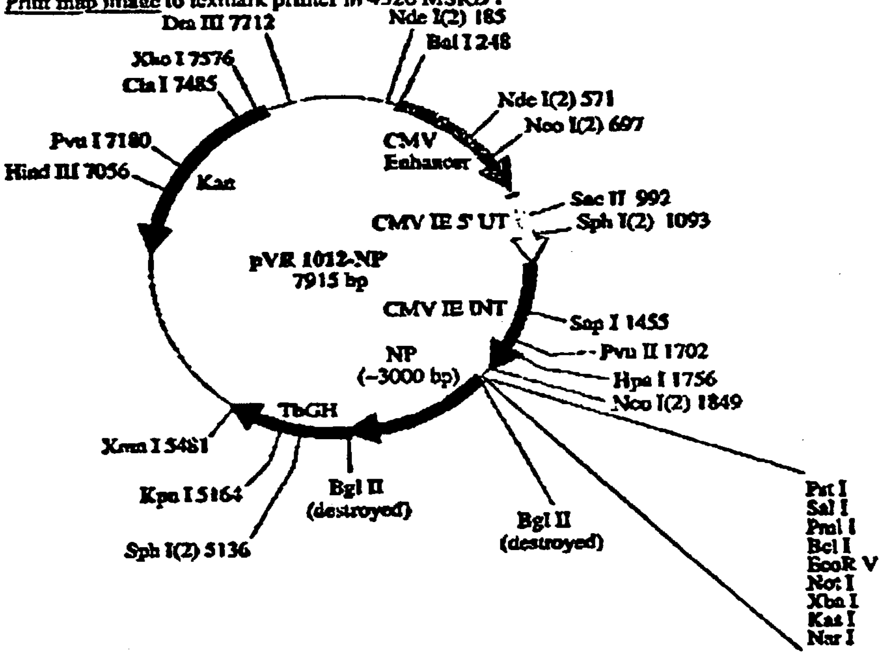

Dra III (7070)
Xho I (6934)
Nde I (185)
Nde I (571)
Nco I (697)
Pvu I (6538)
Hind III (6414)
Sac II (992)
Sph I (1092)
Kan r
CMV enhancer
CMV IE 5' UT
CMV IE INT
pVR 1012-SGP(Z)
7272 bp
Hpa I (1755)
Nco I (1848)
Pml I (1882)
Eco RV (1894)
Not I (1899)
Xba I (1906)
TbGH
SGP(Z)
Kpn I (4522)
Sph I (4494)
Eco RV (2728)

FIGURE 8

Number: 699 **Name:** VR1012-NP **Lab member:** Ling
Backbone origin: [unknown] **Constr. date:** [unknown] **Length(bp):** [unknown]
Keywords: [none]
Comments: [none]

No sequence file available online
No MacPlasmap file available online
Print map image to lexmark printer in 4520 MSRB I



Plasmid name: pVR 1012-NP
Plasmid size: 7915 bp
Constructed by: Ling
Construction date: 1994
Comments/References: none

Figure 9

pVR 1012-GP(IC)

Sequence Listing ID No: 1

General Description

DNA pVR 1012-GP(IC)
Local object
Created: 09/14/98 04:17PM
Last Modification Date: ? (no data)
length: 7003 bp
storage type: Basic
form: Circular

Comments

Restriction Map

BglII: 1 site AGATCT TCTAGA

ClaI: 1 site ATCGAT TAGCTA

DraIII: 1 site CACNNNGTG GTGNNNCAC

EcoRV: 1 site GATATC CTATAG

HindIII: 1 site AAGCTT TTCGAA

HpaI: 1 site GTTAAC CAATTG

KasI: 1 site GGCGCC CCGCGG

KpnI: 1 site GGTACC CCATGG

NarI: 1 site GGCGCC CCGCGG

PmlI: 1 site CACGTG GTGCAC

PstI: 1 site CTGCAG GACGTC

PvuI: 1 site CGATCG GCTAGC

SacII: 1 site CCGCGG GGCGCC

SalI: 1 site GTCGAC CAGCTG

XmnI: 1 site GAANNNNTTC CTTNNNNAAG

EcoRI: 2 sites GAATTC CTTAAG

NcoI: 2 sites CCATGG GGTACC

NdeI: 2 sites CATATG GTATAC

SphI: 2 sites GCATGC CGTACG

XhoI: 2 sites CTCGAG GAGCTC

BamHI: 3 sites GGATCC CCTAGG

BclI: 3 sites TGATCA
ACTAGT

Functional Map

CDS (4 signals)

CMV IE 5' UT

Start: 886 End: 1129

CMV IE INT

Start: 1130 End: 1840

TbGH

Start: 4020 End: 4572

Kan r

Start: 6068 End: 6690 (Complementary)

Misc_feature (2 signals)

CMV enhancer

Start: 248 End: 885

GP(IC)

Start: 1870 End: 4019

Annotations

```
1    TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
     AGCGCGCAAA GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC
.....................................................................
51   GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
     CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT CTGTTCGGGC
.....................................................................
101  TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
     AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGACC GAATTGATAC
.....................................................................
                                     NdeI
151  CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
     GCCGTAGTCT CGTCTAACAT GACTCTCACG TGGTATACGC CACACTTTAT
.....................................................................
201  CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA
     GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCTAACC GATAACCGGT
.....................................................................
251  TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
     AACGTATGCA ACATAGGTAT AGTATTATAC ATGTAAATAT AACCGAGTAC
.....................................................................
301  TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
     AGGTTGTAAT GGCGGTACAA CTGTAACTAA TAACTGATCA ATAATTATCA
.....................................................................
351  AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
     TTAGTTAATG CCCCAGTAAT CAAGTATCGG GTATATACCT CAAGGCGCAA
.....................................................................
401  ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG
     TGTATTGAAT GCCATTTACC GGGCGGACCG ACTGGCGGGT TGCTGGGGGC
.....................................................................
451  CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
     GGGTAACTGC AGTTATTACT GCATACAAGG GTATCATTGC GGTTATCCCT
.....................................................................
501  CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
     GAAAGGTAAC TGCAGTTACC CACCTCATAA ATGCCATTTG ACGGGTGAAC
.....................................................................
                          NdeI
551  GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
     CGTCATGTAG TTCACATAGT ATACGGTTCA TGCGGGGGAT AACTGCAGTT
.....................................................................
601  TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
     ACTGCCATTT ACCGGGCGGA CCGTAATACG GGTCATGTAC TGGAATACCC
.....................................................................
                                                    NcoI
651  ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
     TGAAAGGATG AACCGTCATG TAGATGCATA ATCAGTAGCG ATAATGGTAC
.....................................................................
     NcoI
701  GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC
     CACTACGCCA AAACCGTCAT GTAGTTACCC GCACCTATCG CCAAACTGAG
.....................................................................
751  ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT
     TGCCCCTAAA GGTTCAGAGG TGGGGTAACT GCAGTTACCC TCAAACAAAA
.....................................................................
801  GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA
     CCGTGGTTTT AGTTGCCCTG AAAGGTTTTA CAGCATTGTT GAGGCGGGGT
.....................................................................
```

```
 851  TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
      AACTGCGTTT ACCCGCCATC CGCACATGCC ACCCTCCAGA TATATTCGTC

 901  AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT
      TCGAGCAAAT CACTTGGCAG TCTAGCGGAC CTCTGCGGTA GGTGCGACAA

                                            SacII
 951  TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA
      AACTGGAGGT ATCTTCTGTG GCCCTGGCTA GGTCGGAGGC GCCGGCCCTT

1001  CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC
      GCCACGTAAC CTTGCGCCTA AGGGGCACGG TTCTCACTGC ATTCATGGCG

                                           SphI
1051  CTATAGACTC TATAGGCACA CCCCTTTGGC TCTTATGCAT GCTATACTGT
      GATATCTGAG ATATCCGTGT GGGGAAACCG AGAATACGTA CGATATGACA

1101  TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA TAGGTGATGG
      AAAACCGAAC CCCGGATATG TGGGGGCGAA GGAATACGAT ATCCACTACC

1151  TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
      ATATCGAATC GGATATCCAC ACCCAATAAC TGGTAATAAC TGGTGAGGGG

1201  TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA
      ATAACCACTG CTATGAAAGG TAATGATTAG GTATTGTACC GAGAAACGGT

1251  CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGTCCTTC AGAGACTGAC
      GTTGATAGAG ATAACCGATA TACGGTTATG AGACAGGAAG TCTCTGACTG

1301  ACGGACTCTG TATTTTTACA GGATGGGGTC CCATTTATTA TTTACAAATT
      TGCCTGAGAC ATAAAAATGT CCTACCCCAG GGTAAATAAT AAATGTTTAA

1351  CACATATACA ACAACGCCGT CCCCCGTGCC CGCAGTTTTT ATTAAACATA
      GTGTATATGT TGTTGCGGCA GGGGGCACGG GCGTCAAAAA TAATTTGTAT

1401  GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA CATGGGCTCT
      CGCACCCTAG AGGTGCGCTT AGAGCCCATG CACAAGGCCT GTACCCGAGA

1451  TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
      AGAGGCCATC GCCGCCTCGA AGGTGTAGGC TCGGGACCAG GGTACGGAGG

1501  AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG
      TCGCCGAGTA CCAGCGAGCC GTCGAGGAAC GAGGATTGTC ACCTCCGGTC

1551  ACTTAGGCAC AGCACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG
      TGAATCCGTG TCGTGTTACG GGTGGTGGTG GTCACACGGC GTGTTCCGGC

1601  TGGCGGTAGG GTATGTGTCT GAAAATGAGC GTGGAGATTG GGCTCGCACG
      ACCGCCATCC CATACACAGA CTTTTACTCG CACCTCTAAC CCGAGCGTGC

1651  GCTGACGCAG ATGGAAGACT TAAGGCAGCG GCAGAAGAAG ATGCAGGCAG
      CGACTGCGTC TACCTTCTGA ATTCCGTCGC CGTCTTCTTC TACGTCCGTC

1701  CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC GTTGCGGTGC
      GACTCAACAA CATAAGACTA TTCTCAGTCT CCATTGAGGG CAACGCCACG
```

```
       HpaI
1751  TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
      ACAATTGCCA CCTCCCGTCA CATCAGACTC GTCATGAGCA ACGACGGCGC

                                                    NcoI
1801  CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT
      GCGCGGTGGT CTGTATTATC GACTGTCTGA TTGTCTGACA AGGAAAGGTA

                                SalI
      NcoI       PstI                  PmlI   BclI    EcoRV
1851  GGGTCTTTTC TGCAGTCACC GTCGTCGACA CGTGTGATCA GATATCGCGG
      CCCAGAAAAG ACGTCAGTGG CAGCAGCTGT GCACACTAGT CTATAGCGCC

                      EcoRI
1901  CCGCGCGGCC GCTCTAGAAT TCTCTAATCA CAGTCATCAT GGGAGCGTCA
      GGCGCGCCGG CGAGATCTTA AGAGATTAGT GTCAGTAGTA CCCTCGCAGT

1951  GGGATTCTGC AATTGCCCCG TGAGCGCTTC AGGAAAACAT CTTTCTTTGT
      CCCTAAGACG TTAACGGGGC ACTCGCGAAG TCCTTTTGTA GAAAGAAACA

2001  TTGGGTAATA ATCCTATTCC ATAAAGTCTT TTCAATCCCG TTGGGGGTTG
      AACCCATTAT TAGGATAAGG TATTTCAGAA AAGTTAGGGC AACCCCCAAC

2051  TACACAACAA TACCCTACAA GTGAGTGATA TTGACAAGTT TGTGTGCCGA
      ATGTGTTGTT ATGGGATGTT CACTCACTAT AACTGTTCAA ACACACGGCT

2101  GACAAACTCT CTTCAACTAG CCAATTGAAG TCAGTCGGGT TGAACTTGGA
      CTGTTTGAGA GAAGTTGATC GGTTAACTTC AGTCAGCCCA ACTTGAACCT

2151  GGGCAATGGA GTAGCAACTG ATGTACCAAC GGCAACCAAA AGATGGGGTT
      CCCGTTACCT CATCGTTGAC TACATGGTTG CCGTTGGTTT TCTACCCCAA

2201  TTCGAGCTGG TGTTCCACCA AAGGTGGTAA ATTACGAAGC TGGAGAATGG
      AAGCTCGACC ACAAGGTGGT TTCCACCATT TAATGCTTCG ACCTCTTACC

2251  GCTGAGAACT GTTATAACCT GGCTATAAAG AAAGTTGATG GTAGTGAGTG
      CGACTCTTGA CAATATTGGA CCGATATTTC TTTCAACTAC CATCACTCAC

2301  CCTACCAGAA GCCCCTGAGG GAGTGAGGGA TTTTCCCCGT TGCCGCTATG
      GGATGGTCTT CGGGGACTCC CTCACTCCCT AAAAGGGGCA ACGGCGATAC

2351  TACACAAAGT CTCAGGAACT GGACCATGCC CAGGAGGACT CGCCTTTCAC
      ATGTGTTTCA GAGTCCTTGA CCTGGTACGG GTCCTCCTGA GCGGAAAGTG

2401  AAAGAAGGAG CCTTCTTCCT GTATGACCGA CTCGCATCAA CAATCATTTA
      TTTCTTCCTC GGAAGAAGGA CATACTGGCT GAGCGTAGTT GTTAGTAAAT

2451  TCGGGGTACA ACCTTTGCCG AAGGAGTTAT TGCATTTCTG ATCTTGCCTA
      AGCCCCATGT TGGAAACGGC TTCCTCAATA ACGTAAAGAC TAGAACGGAT

2501  AGGCGCGAAA GGATTTTTTC CAGTCTCCTC CATTGCATGA GCCTGCCAAC
      TCCGCGCTTT CCTAAAAAAG GTCAGAGGAG GTAACGTACT CGGACGGTTG
```

```
            BamHI
2551  ATGACCACGG ATCCCTCCAG TTACTATCAC ACGACAACAA TAAACTACGT
      TACTGGTGCC TAGGGAGGTC AATGATAGTG TGCTGTTGTT ATTTGATGCA

2601  GGTTGATAAT TTTGGAACCA ACACCACAGA GTTTCTGTTC CAAGTCGATC
      CCAACTATTA AAACCTTGGT TGTGGTGTCT CAAAGACAAG GTTCAGCTAG

                       XhoI
2651  ATTTGACGTA TGTGCAGCTC GAGGCAAGAT TCACACCACA ATTCCTTGTC
      TAAACTGCAT ACACGTCGAG CTCCGTTCTA AGTGTGGTGT TAAGGAACAG

2701  CTCCTAAATG AAACCATCTA CTCTGATAAC CGCAGAAGTA ACACAACAGG
      GAGGATTTAC TTTGGTAGAT GAGACTATTG GCGTCTTCAT TGTGTTGTCC

2751  AAAACTAATC TGGAAAATAA ATCCCACTGT TGATACCAGC ATGGGTGAGT
      TTTTGATTAG ACCTTTTATT TAGGGTGACA ACTATGGTCG TACCCACTCA

2801  GGGCTTTCTG GGAAAATAAA AAAACTTCAC AAAAACCCTT TCAAGTGAAG
      CCCGAAAGAC CCTTTTATTT TTTTGAAGTG TTTTTGGGAA AGTTCACTTC

2851  AGTTGTCTTT CGTACCTGTA CCAGAAACCC AGAACCAGGT CCTTGACACG
      TCAACAGAAA GCATGGACAT GGTCTTTGGG TCTTGGTCCA GGAACTGTGC

2901  ACAGCGACGG TCTCTCCTCC CATCTCCGCC CACAACCACG CAGGCGAAGA
      TGTCGCTGCC AGAGAGGAGG GTAGAGGCGG GTGTTGGTGC GTCCGCTTCT

2951  CCACAAAGAA TTGGTTTCAG AGGATTCCAC TCCAGTGGTT CAGATGCAAA
      GGTGTTTCTT AACCAAAGTC TCCTAAGGTG AGGTCACCAA GTCTACGTTT

3001  ACATCAAGGG AAAGGACACA ATGCCAACCA CAGTGACGGG TGTACCAACA
      TGTAGTTCCC TTTCCTGTGT TACGGTTGGT GTCACTGCCC ACATGGTTGT

                                          BclI
3051  ACCACACCCT CTCCATTTCC AATCAATGCT CGCAACACTG ATCATACCAA
      TGGTGTGGGA GAGGTAAAGG TTAGTTACGA GCGTTGTGAC TAGTATGGTT

3101  ATCATTTATC GGCCTGGAGG GGCCCCAAGA AGACCACAGC ACCACACAGC
      TAGTAAATAG CCGGACCTCC CCGGGGTTCT TCTGGTGTCG TGGTGTGTCG

3151  CTGCCAAGAC CACCAGCCAA CCAACCAACA GCACAGAATC GACGACACTA
      GACGGTTCTG GTGGTCGGTT GGTTGGTTGT CGTGTCTTAG CTGCTGTGAT

3201  AACCCAACAT CAGAGCCCTC CAGTAGAGGC ACGGGACCAT CCAGCCCCAC
      TTGGGTTGTA GTCTCGGGAG GTCATCTCCG TGCCCTGGTA GGTCGGGGTG

3251  GGTCCCCAAC ACCACAGAAA GCCACGCCGA ACTTGGCAAG ACAACCCCAA
      CCAGGGGTTG TGGTGTCTTT CGGTGCGGCT TGAACCGTTC TGTTGGGGTT

3301  CCACACTCCC AGAACAGCAC ACTGCCGCCA GTGCCATTCC AAGAGCCGTG
      GGTGTGAGGG TCTTGTCGTG TGACGGCGGT CACGGTAAGG TTCTCGGCAC

3351  CACCCCGACG AACTCAGTGG ACCTGGCTTC CTGACGAACA CAATACGGGG
      GTGGGGCTGC TTGAGTCACC TGGACCGAAG GACTGCTTGT GTTATGCCCC
```

```
                              BamHI
                              ------
3401  GGTGACAAAT CTCCTGACAG GATCCAGAAG AAAGCGAAGG GATGTCACTC
      CCACTGTTTA GAGGACTGTC CTAGGTCTTC TTTCGCTTCC CTACAGTGAG
...........................................................
3451  CCAATACACA ACCCAAATGC AACCCAAACC TGCACTATTG GACAGCCTTG
      GGTTATGTGT TGGGTTTACG TTGGGTTTGG ACGTGATAAC CTGTCGGAAC
...........................................................
3501  GATGAGGGTG CTGCCATAGG TTTAGCCTGG ATACCATACT TCGGGCCAGC
      CTACTCCCAC GACGGTATCC AAATCGGACC TATGGTATGA AGCCCGGTCG
...........................................................
3551  AGCTGAGGGA ATTTACACTG AAGGCATAAT GGAGAATCAA AATGGATTGA
      TCGACTCCCT TAAATGTGAC TTCCGTATTA CCTCTTAGTT TTACCTAACT
...........................................................
3601  TCTGTGGATT GAGGCAGCTG GCCAACGAAA CGACACAAGC TCTTCAATTG
      AGACACCTAA CTCCGTCGAC CGGTTGCTTT GCTGTGTTCG AGAAGTTAAC
...........................................................
3651  TTCTTAAGGG CAACTACTGA GTTGCGTACA TTCTCTATAC TAAATCGGAA
      AAGAATTCCC GTTGATGACT CAACGCATGT AAGAGATATG ATTTAGCCTT
...........................................................
3701  AGCAATAGAC TTCTTGCTCC AAAGATGGGG AGGAACATGT CACATTCTAG
      TCGTTATCTG AAGAACGAGG TTTCTACCCC TCCTTGTACA GTGTAAGATC
...........................................................
3751  GGCCTGATTG TTGCATTGAA CCCCAAGATT GGACCAAAAA TATCACTGAT
      CCGGACTAAC AACGTAACTT GGGGTTCTAA CCTGGTTTTT ATAGTGACTA
...........................................................
         BclI
         ------
3801  AAAATTGATC AAATAATCCA TGACTTTGTC GATAATAATC TTCCAAATCA
      TTTTAACTAG TTTATTAGGT ACTGAAACAG CTATTATTAG AAGGTTTAGT
...........................................................
3851  GAATGATGGC AGCAACTGGT GGACTGGATG GAAACAATGG GTTCCTGCTG
      CTTACTACCG TCGTTGACCA CCTGACCTAC CTTTGTTACC CAAGGACGAC
...........................................................
3901  GAATAGGAAT CACAGGAGTA ATCATTGCTA TTATTGCTTT GCTGTGCATT
      CTTATCCTTA GTGTCCTCAT TAGTAACGAT AATAACGAAA CGACACGTAA
...........................................................
                                                 EcoRI
                                                 ------
3951  TGCAAATTCA TGCTTTGAAC TAATATAGCA TCATACTTTA GAATTCTAGA
      ACGTTTAAGT ACGAAACTTG ATTATATCGT AGTATGAAAT CTTAAGATCT
...........................................................
        NarI
        ------
        KasI    BamHI BglII
        ------  -----------
4001  CCAGGCGCCT GGATCCAGAT CTGCTGTGCC TTCTAGTTGC CAGCCATCTG
      GGTCCGCGGA CCTAGGTCTA GACGACACGG AAGATCAACG GTCGGTAGAC
...........................................................
4051  TTGTTTGCCC CTCCCCCGTG CCTTCCTTGA CCCTGGAAGG TGCCACTCCC
      AACAAACGGG GAGGGGGCAC GGAAGGAACT GGGACCTTCC ACGGTGAGGG
...........................................................
4101  ACTGTCCTTT CCTAATAAAA TGAGGAAATT GCATCGCATT GTCTGAGTAG
      TGACAGGAAA GGATTATTTT ACTCCTTTAA CGTAGCGTAA CAGACTCATC
...........................................................
4151  GTGTCATTCT ATTCTGGGGG GTGGGGTGGG GCAGCACAGC AAGGGGGAGG
      CACAGTAAGA TAAGACCCCC CACCCCACCC CGTCGTGTCG TTCCCCCTCC
...........................................................
```

```
                                SphI                                     KpnI
4201  ATTGGGAAGA CAATAGCAGG CATGCTGGGG ATGCGGTGGG CTCTATGGGT
      TAACCCTTCT GTTATCGTCC GTACGACCCC TACGCCACCC GAGATACCCA
      ...........................................................................
      KpnI
4251  ACCCAGGTGC TGAAGAATTG ACCCGGTTCC TCCTGGGCCA GAAAGAAGCA
      TGGGTCCACG ACTTCTTAAC TGGGCCAAGG AGGACCCGGT CTTTCTTCGT
      ...........................................................................
4301  GGCACATCCC CTTCTCTGTG ACACACCCTG TCCACGCCCC TGGTTCTTAG
      CCGTGTAGGG GAAGAGACAC TGTGTGGGAC AGGTGCGGGG ACCAAGAATC
      ...........................................................................
4351  TTCCAGCCCC ACTCATAGGA CACTCATAGC TCAGGAGGGC TCCGCCTTCA
      AAGGTCGGGG TGAGTATCCT GTGAGTATCG AGTCCTCCCG AGGCGGAAGT
      ...........................................................................
4401  ATCCCACCCG CTAAAGTACT TGGAGCGGTC TCTCCCTCCC TCATCAGCCC
      TAGGGTGGGC GATTTCATGA ACCTCGCCAG AGAGGGAGGG AGTAGTCGGG
      ...........................................................................
4451  ACCAAACCAA ACCTAGCCTC CAAGAGTGGG AAGAAATTAA AGCAAGATAG
      TGGTTTGGTT TGGATCGGAG GTTCTCACCC TTCTTTAATT TCGTTCTATC
      ...........................................................................
4501  GCTATTAAGT GCAGAGGGAG AGAAAATGCC TCCAACATGT GAGGAAGTAA
      CGATAATTCA CGTCTCCCTC TCTTTTACGG AGGTTGTACA CTCCTTCATT
      ...........................................................................
                           XmnI
4551  TGAGAGAAAT CATAGAATTT CTTCCGCTTC CTCGCTCACT GACTCGCTGC
      ACTCTCTTTA GTATCTTAAA GAAGGCGAAG GAGCGAGTGA CTGAGCGACG
      ...........................................................................
4601  GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA
      CGAGCCAGCA AGCCGACGCC GCTCGCCATA GTCGAGTGAG TTTCCGCCAT
      ...........................................................................
4651  ATACGGTTAT CCACAGAATC AGGGGATAAC GCAGGAAAGA ACATGTGAGC
      TATGCCAATA GGTGTCTTAG TCCCCTATTG CGTCCTTTCT TGTACACTCG
      ...........................................................................
4701  AAAAGGCCAG CAAAAGGCCA GGAACCGTAA AAAGGCCGCG TTGCTGGCGT
      TTTTCCGGTC GTTTTCCGGT CCTTGGCATT TTTCCGGCGC AACGACCGCA
      ...........................................................................
4751  TTTTCCATAG GCTCCGCCCC CCTGACGAGC ATCACAAAAA TCGACGCTCA
      AAAAGGTATC CGAGGCGGGG GGACTGCTCG TAGTGTTTTT AGCTGCGAGT
      ...........................................................................
4801  AGTCAGAGGT GGCGAAACCC GACAGGACTA TAAAGATACC AGGCGTTTCC
      TCAGTCTCCA CCGCTTTGGG CTGTCCTGAT ATTTCTATGG TCCGCAAAGG
      ...........................................................................
4851  CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG
      GGGACCTTCG AGGGAGCACG CGAGAGGACA AGGCTGGGAC GGCGAATGGC
      ...........................................................................
4901  GATACCTGTC CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCAATGC
      CTATGGACAG GCGGAAAGAG GGAAGCCCTT CGCACCGCGA AAGAGTTACG
      ...........................................................................
4951  TCACGCTGTA GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG
      AGTGCGACAT CCATAGAGTC AAGCCACATC CAGCAAGCGA GGTTCGACCC
      ...........................................................................
5001  CTGTGTGCAC GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA
      GACACACGTG CTTGGGGGGC AAGTCGGGCT GGCGACGCGG AATAGGCCAT
      ...........................................................................
```

5051 ACTATCGTCT TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA
TGATAGCAGA ACTCAGGTTG GGCCATTCTG TGCTGAATAG CGGTGACCGT

5101 GCAGCCACTG GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC
CGTCGGTGAC CATTGTCCTA ATCGTCTCGC TCCATACATC CGCCACGATG

5151 AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGGACAGTAT
TCTCAAGAAC TTCACCACCG GATTGATGCC GATGTGATCT TCCTGTCATA

5201 TTGGTATCTG CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT
AACCATAGAC GCGAGACGAC TTCGGTCAAT GGAAGCCTTT TTCTCAACCA

5251 AGCTCTTGAT CCGGCAAACA AACCACCGCT GGTAGCGGTG GTTTTTTTGT
TCGAGAACTA GGCCGTTTGT TTGGTGGCGA CCATCGCCAC CAAAAAAACA

5301 TTGCAAGCAG CAGATTACGC GCAGAAAAAA AGGATCTCAA GAAGATCCTT
AACGTTCGTC GTCTAATGCG CGTCTTTTTT TCCTAGAGTT CTTCTAGGAA

5351 TGATCTTTTC TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTTAA
ACTAGAAAAG ATGCCCCAGA CTGCGAGTCA CCTTGCTTTT GAGTGCAATT

5401 GGGATTTTGG TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT
CCCTAAAACC AGTACTCTAA TAGTTTTTCC TAGAAGTGGA TCTAGGAAAA

5451 AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT
TTTAATTTTT ACTTCAAAAT TTAGTTAGAT TTCATATATA CTCATTTGAA

5501 GGTCTGACAG TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC
CCAGACTGTC AATGGTTACG AATTAGTCAC TCCGTGGATA GAGTCGCTAG

5551 TGTCTATTTC GTTCATCCAT AGTTGCCTGA CTCCGGGGGG GGGGGGCGCT
ACAGATAAAG CAAGTAGGTA TCAACGGACT GAGGCCCCCC CCCCCCGCGA

5601 GAGGTCTGCC TCGTGAAGAA GGTGTTGCTG ACTCATACCA GGCCTGAATC
CTCCAGACGG AGCACTTCTT CCACAACGAC TGAGTATGGT CCGGACTTAG

5651 GCCCCATCAT CCAGCCAGAA AGTGAGGGAG CCACGGTTGA TGAGAGCTTT
CGGGGTAGTA GGTCGGTCTT TCACTCCCTC GGTGCCAACT ACTCTCGAAA

5701 GTTGTAGGTG GACCAGTTGG TGATTTTGAA CTTTTGCTTT GCCACGGAAC
CAACATCCAC CTGGTCAACC ACTAAAACTT GAAAACGAAA CGGTGCCTTG

5751 GGTCTGCGTT GTCGGGAAGA TGCGTGATCT GATCCTTCAA CTCAGCAAAA
CCAGACGCAA CAGCCCTTCT ACGCACTAGA CTAGGAAGTT GAGTCGTTTT

5801 GTTCGATTTA TTCAACAAAG CCGCCGTCCC GTCAAGTCAG CGTAATGCTC
CAACCTAAAT AAGTTGTTTC GGCGGCAGGG CAGTTCAGTC GCATTACGAG

5851 TGCCAGTGTT ACAACCAATT AACCAATTCT GATTAGAAAA ACTCATCGAG
ACGGTCACAA TGTTGGTTAA TTGGTTAAGA CTAATCTTTT TGAGTAGCTC

5901 CATCAAATGA AACTGCAATT TATTCATATC AGGATTATCA ATACCATATT
GTAGTTTACT TTGACGTTAA ATAAGTATAG TCCTAATAGT TATGGTATAA

5951 TTTGAAAAAG CCGTTTCTGT AATGAAGGAG AAAACTCACC GAGGCAGTTC
AAACTTTTTC GGCAAAGACA TTACTTCCTC TTTTGAGTGG CTCCGTCAAG

```
6001 CATAGGATGG CAAGATCCTG GTATCGGTCT GCGATTCCGA CTCGTCCAAC
     GTATCCTACC GTTCTAGGAC CATAGCCAGA CGCTAAGGCT GAGCAGGTTG
.......................................................................
6051 ATCAATACAA CCTATTAATT TCCCCTCGTC AAAAATAAGG TTATCAAGTG
     TAGTTATGTT GGATAATTAA AGGGGAGCAG TTTTTATTCC AATAGTTCAC
.......................................................................
                                                 HindIII
6101 AGAAATCACC ATGAGTGACG ACTGAATCCG GTGAGAATGG CAAAAGCTTA
     TCTTTAGTGG TACTCACTGC TGACTTAGGC CACTCTTACC GTTTTCGAAT
.......................................................................
6151 TGCATTTCTT TCCAGACTTG TTCAACAGGC CAGCCATTAC GCTCGTCATC
     ACGTAAAGAA AGGTCTGAAC AAGTTGTCCG GTCGGTAATG CGAGCAGTAG
.......................................................................
6201 AAAATCACTC GCATCAACCA AACCGTTATT CATTCGTGAT TGCGCCTGAG
     TTTTAGTGAG CGTAGTTGGT TTGGCAATAA GTAAGCACTA ACGCGGACTC
.......................................................................
                PvuI
6251 CGAGACGAAA TACGCGATCG CTGTTAAAAG GACAATTACA AACAGGAATC
     GCTCTGCTTT ATGCGCTAGC GACAATTTTC CTGTTAATGT TTGTCCTTAG
.......................................................................
6301 GAATGCAACC GGCGCAGGAA CACTGCCAGC GCATCAACAA TATTTTCACC
     CTTACGTTGG CCGCGTCCTT GTGACGGTCG CGTAGTTGTT ATAAAAGTGG
.......................................................................
6351 TGAATCAGGA TATTCTTCTA ATACCTGGAA TGCTGTTTTC CCGGGGATCG
     ACTTAGTCCT ATAAGAAGAT TATGGACCTT ACGACAAAAG GGCCCCTAGC
.......................................................................
6401 CAGTGGTGAG TAACCATGCA TCATCAGGAG TACGGATAAA ATGCTTGATG
     GTCACCACTC ATTGGTACGT AGTAGTCCTC ATGCCTATTT TACGAACTAC
.......................................................................
6451 GTCGGAAGAG GCATAAATTC CGTCAGCCAG TTTAGTCTGA CCATCTCATC
     CAGCCTTCTC CGTATTTAAG GCAGTCGGTC AAATCAGACT GGTAGAGTAG
.......................................................................
6501 TGTAACATCA TTGGCAACGC TACCTTTGCC ATGTTTCAGA AACAACTCTG
     ACATTGTAGT AACCGTTGCG ATGGAAACGG TACAAAGTCT TTGTTGAGAC
.......................................................................
                           ClaI
6551 GCGCATCGGG CTTCCCATAC AATCGATAGA TTGTCGCACC TGATTGCCCG
     CGCGTAGCCC GAAGGGTATG TTAGCTATCT AACAGCGTGG ACTAACGGGC
.......................................................................
6601 ACATTATCGC GAGCCCATTT ATACCCATAT AAATCAGCAT CCATGTTGGA
     TGTAATAGCG CTCGGGTAAA TATGGGTATA TTTAGTCGTA GGTACAACCT
.......................................................................
                XhoI
6651 ATTTAATCGC GGCCTCGAGC AAGACGTTTC CCGTTGAATA TGGCTCATAA
     TAAATTAGCG CCGGAGCTCG TTCTGCAAAG GGCAACTTAT ACCGAGTATT
.......................................................................
6701 CACCCCTTGT ATTACTGTTT ATGTAAGCAG ACAGTTTTAT TGTTCATGAT
     GTGGGGAACA TAATGACAAA TACATTCGTC TGTCAAAATA ACAAGTACTA
.......................................................................
                                                 DraIII
6751 GATATATTTT TATCTTGTGC AATGTAACAT CAGAGATTTT GAGACACAAC
     CTATATAAAA ATAGAACACG TTACATTGTA GTCTCTAAAA CTCTGTGTTG
.......................................................................
```

DraIII

6801 GTGGCTTTCC CCCCCCCCCC ATTATTGAAG CATTTATCAG GGTTATTGTC
CACCGAAAGG GGGGGGGGGG TAATAACTTC GTAAATAGTC CCAATAACAG

6851 TCATGAGCGG ATACATATTT GAATGTATTT AGAAAAATAA ACAAATAGGG
AGTACTCGCC TATGTATAAA CTTACATAAA TCTTTTTATT TGTTTATCCC

6901 GTTCCGCGCA CATTTCCCCG AAAAGTGCCA CCTGACGTCT AAGAAACCAT
CAAGGCGCGT GTAAAGGGGC TTTTCACGGT GGACTGCAGA TTCTTTGGTA

6951 TATTATCATG ACATTAACCT ATAAAAATAG GCGTATCACG AGGCCCTTTC
ATAATAGTAC TGTAATTGGA TATTTTTATC CGCATAGTGC TCCGGGAAAG

7001 GTC
CAG

pVR 1012-GP(S)

Sequence Listing ID No: 2

General Description

DNA pVR 1012-GP(S)
Local object
Created: 09/14/98 03:58PM
Last Modification Date: ? (no data)
length: 7073 bp
storage type: Basic
form: Circular

Comments

Restriction Map

BalI: 1 site	TGGCCA ACCGGT
BclI: 1 site	TGATCA ACTAGT
ClaI: 1 site	ATCGAT TAGCTA
DraIII: 1 site	CACNNNGTG GTGNNNCAC
HindIII: 1 site	AAGCTT TTCGAA
KasI: 1 site	GGCGCC CCGCGG
KpnI: 1 site	GGTACC CCATGG
NarI: 1 site	GGCGCC CCGCGG
PmlI: 1 site	CACGTG GTGCAC
PvuI: 1 site	CGATCG GCTAGC
SacII: 1 site	CCGCGG GGCGCC
SalI: 1 site	GTCGAC CAGCTG
XbaI: 1 site	TCTAGA AGATCT
XmnI: 1 site	GAANNNNTTC CTTNNNNAAG
NdeI: 2 sites	CATATG GTATAC
EcoRV: 3 sites	GATATC CTATAG
SphI: 3 sites	GCATGC CGTACG
NcoI: 4 sites	CCATGG GGTACC
BamHI: 6 sites	GGATCC CCTAGG

Functional Map

CDS (4 signals)

CMV IE 5' UT

Start: 886 End: 1129

CMV IE INT

Start: 1130 End: 1840

TbGH

Start: 4090 End: 4642

Kan r

Start: 6138 End: 6760 (Complementary)

Misc_feature (2 signals)

CMV enhancer

Start: 248 End: 885

GP(S)

Start: 1870 End: 4089

Annotations

```
  1  TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
     AGCGCGCAAA GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC

 51  GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
     CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT CTGTTCGGGC

101  TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
     AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGACC GAATTGATAC

                                       NdeI
151  CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
     GCCGTAGTCT CGTCTAACAT GACTCTCACG TGGTATACGC CACACTTTAT

                                                 BalI
201  CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA
     GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCTAACC GATAACCGGT

251  TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
     AACGTATGCA ACATAGGTAT AGTATTATAC ATGTAAATAT AACCGAGTAC

301  TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
     AGGTTGTAAT GGCGGTACAA CTGTAACTAA TAACTGATCA ATAATTATCA

351  AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
     TTAGTTAATG CCCCAGTAAT CAAGTATCGG GTATATACCT CAAGGCGCAA

401  ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG
     TGTATTGAAT GCCATTTACC GGGCGGACCG ACTGGCGGGT TGCTGGGGGC

451  CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
     GGGTAACTGC AGTTATTACT GCATACAAGG GTATCATTGC GGTTATCCCT

501  CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
     GAAAGGTAAC TGCAGTTACC CACCTCATAA ATGCCATTTG ACGGGTGAAC

                            NdeI
551  GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
     CGTCATGTAG TTCACATAGT ATACGGTTCA TGCGGGGGAT AACTGCAGTT

601  TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
     ACTGCCATTT ACCGGGCGGA CCGTAATACG GGTCATGTAC TGGAATACCC

                                                 NcoI
651  ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
     TGAAAGCATG AACCGTCATG TAGATGCATA ATCAGTAGCG ATAATGGTAC

     NcoI
701  GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC
     CACTACGCCA AAACCGTCAT GTAGTTACCC GCACCTATCG CCAAACTGAG

751  ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT
     TGCCCCTAAA GGTTCAGAGG TGGGGTAACT GCAGTTACCC TCAAACAAAA
```

```
801  GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA
     CCGTGGTTTT AGTTGCCCTG AAAGGTTTTA CAGCATTGTT GAGGCGGGGT

851  TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
     AACTGCGTTT ACCCGCCATC CGCACATGCC ACCCTCCAGA TATATTCGTC

901  AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT
     TCGAGCAAAT CACTTGGCAG TCTAGCGGAC CTCTGCGGTA GGTGCGACAA

                                                SacII
951  TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA
     AACTGGAGGT ATCTTCTGTG GCCCTGGCTA GGTCGGAGGC GCCGGCCCTT

1001 CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC
     GCCACGTAAC CTTGCGCCTA AGGGGCACGG TTCTCACTGC ATTCATGGCG

                                               SphI
1051 CTATAGACTC TATAGGCACA CCCCTTTGGC TCTTATGCAT GCTATACTGT
     GATATCTGAG ATATCCGTGT GGGGAAACCG AGAATACGTA CGATATGACA

1101 TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA TAGGTGATGG
     AAAACCGAAC CCCGGATATG TGGGGGCGAA GGAATACGAT ATCCACTACC

1151 TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
     ATATCGAATC GGATATCCAC ACCCAATAAC TGGTAATAAC TGGTGAGGGG

1201 TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA
     ATAACCACTG CTATGAAAGG TAATGATTAG GTATTGTACC GAGAAACGGT

1251 CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGTCCTTC AGAGACTGAC
     GTTGATAGAG ATAACCGATA TACGGTTATG AGACAGGAAG TCTCTGACTG

1301 ACGGACTCTG TATTTTTACA GGATGGGGTC CCATTTATTA TTTACAAATT
     TGCCTGAGAC ATAAAAATGT CCTACCCCAG GGTAAATAAT AAATGTTTAA

1351 CACATATACA ACAACGCCGT CCCCCGTGCC CGCAGTTTTT ATTAAACATA
     GTGTATATGT TGTTGCGGCA GGGGGCACGG GCGTCAAAAA TAATTTGTAT

1401 GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA CATGGGCTCT
     CGCACCCTAG AGGTGCGCTT AGAGCCCATG CACAAGGCCT GTACCCGAGA

1451 TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
     AGAGGCCATC GCCGCCTCGA AGGTGTAGGC TCGGGACCAG GGTACGGAGG

1501 AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG
     TCGCCGAGTA CCAGCGAGCC GTCGAGGAAC GAGGATTGTC ACCTCCGGTC

1551 ACTTAGGCAC AGCACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG
     TGAATCCGTG TCGTGTTACG GGTGGTGGTG GTCACACGGC GTGTTCCGGC

1601 TGGCGGTAGG GTATGTGTCT GAAAATGAGC GTGGAGATTG GGCTCGCACG
     ACCGCCATCC CATACACAGA CTTTTACTCG CACCTCTAAC CCGAGCGTGC

1651 GCTGACGCAG ATGGAAGACT TAAGGCAGCG GCAGAAGAAG ATGCAGGCAG
     CGACTGCGTC TACCTTCTGA ATTCCGTCGC CGTCTTCTTC TACGTCCGTC
```

```
1701  CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC GTTGCCGTGC
      GACTCAACAA CATAAGACTA TTCTCAGTCT CCATTGAGGG CAACGCCACG
...........................................................

1751  TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
      ACAATTGCCA CCTCCCGTCA CATCAGACTC GTCATGAGCA ACGACGGCGC
...........................................................

                                                    NcoI
1801  CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT
      GCGCGGTGGT CTGTATTATC GACTGTCTGA TTGTCTGACA AGGAAAGGTA
...........................................................

                            SalI
      NcoI                      PmlI    BclI    EcoRV
1851  GGGTCTTTTC TGCAGTCACC GTCGTCGACA CGTGTGATCA GATATCGCGG
      CCCAGAAAAG ACGTCAGTGG CAGCAGCTGT GCACACTAGT CTATAGCGCC
...........................................................

                     SphI                          EcoRV
1901  CCGCTCTAGC TAGATGCATG CTCGAGCGGC CGCCAGTGTG ATGGATATCT
      GGCGAGATCG ATCTACGTAC GAGCTCGCCG GCGGTCACAC TACCTATAGA
...........................................................

                                NcoI
1951  GCAGAATTCT ATCTTCAGGA TCTCGCCATG GAGGGTCTTA GCCTACTCCA
      CGTCTTAAGA TAGAAGTCCT AGAGCGGTAC CTCCCAGAAT CGGATGAGGT
...........................................................
2001  ATTGCCCAGA GATAAATTTC GAAAAAGCTC TTTCTTTGTT TGGGTCATCA
      TAACGGGTCT CTATTTAAAG CTTTTTCGAG AAAGAAACAA ACCCAGTAGT
...........................................................
2051  TCTTATTTCA AAAGGCCTTT TCCATGCCTT TGGGTGTTGT GACCAACAGC
      AGAATAAAGT TTTCCGGAAA AGGTACGGAA ACCCACAACA CTGGTTGTCG
...........................................................
2101  ACTTTAGAAG TAACAGAGAT TGACCAGCTA GTCTGCAAGG ATCATCTTGC
      TGAAATCTTC ATTGTCTCTA ACTGGTCGAT CAGACGTTCC TAGTAGAACG
...........................................................
2151  ATCAACTGAC CAGCTGAAAT CAGTTGGTCT CAACCTCGAG GGGAGCGGAG
      TAGTTGACTG GTCGACTTTA GTCAACCAGA GTTGGAGCTC CCCTCGCCTC
...........................................................

           EcoRV
2201  TATCTACTGA TATCCCATCT GCGACAAAGC GTTGGGGCTT CAGATCTGGT
      ATAGATGACT ATAGGGTAGA CGCTGTTTCG CAACCCCGAA GTCTAGACCA
...........................................................
2251  GTGCCTCCCC AAGTGGTCAG CTATGAAGCA GGAGAATGGG CTGAAAATTG
      CACGGAGGGG TTCACCAGTC GATACTTCGT CCTCTTACCC GACTTTTAAC
...........................................................
2301  CTACAATCTT GAAATAAAGA AACCGGACGG GAGCGAATGC TTACCCCCAC
      GATGTTAGAA CTTTATTTCT TTGGCCTGCC CTCGCTTACG AATGGGGGTG
...........................................................
2351  CGCCGGATGG TGTCAGAGGC TTTCCAAGGT GCCGCTATGT TCACAAAGCC
      GCGGCCTACC ACAGTCTCCG AAAGGTTCCA CGGCGATACA AGTGTTTCGG
...........................................................
2401  CAAGGAACCG GGCCCTGCCC GGGTGACTAT GCCTTTCACA AGGATGGAGC
      GTTCCTTGGC CCGGGACGGG CCCACTGATA CGGAAAGTGT TCCTACCTCG
...........................................................
2451  TTTCTTCCTC TATGACAGGC TGGCTTCAAC TGTAATTTAC AGAGGAGTCA
      AAAGAAGGAG ATACTGTCCG ACCGAAGTTG ACATTAAATG TCTCCTCAGT
...........................................................
```

```
2501 ATTTTGCTGA GGGGGTAATC GCATTCTTGA TATTGGCTAA ACCAAAGGAA
     TAAAACGACT CCCCCATTAG CGTAAGAACT ATAACCGATT TGGTTTCCTT

2551 ACGTTCCTTC AATCACCCCC CATTCGAGAG GCAGCAAACT ACACTGAAAA
     TGCAAGGAAG TTAGTGGGGG GTAAGCTCTC CGTCGTTTGA TGTGACTTTT

2601 TACATCAAGT TACTATGCCA CATCCTACTT GGAGTACGAA ATCGAAAATT
     ATGTAGTTCA ATGATACGGT GTAGGATGAA CCTCATGCTT TAGCTTTTAA

2651 TTGGTGCTCA ACACTCCACG ACCCTTTTCA AAATTAACAA TAATACTTTT
     AACCACGAGT TGTGAGGTGC TGGGAAAAGT TTTAATTGTT ATTATGAAAA

2701 GTTCTTCTGG ACAGGCCCCA CACGCCTCAG TTCCTTTTCC AGCTGAATGA
     CAAGAAGACC TGTCCGGGGT GTGCGGAGTC AAGGAAAAGG TCGACTTACT

2751 TACCATTCAA CTTCACCAAC AGTTGAGCAA CACAACTGGG AAACTAATTT
     ATGGTAAGTT GAAGTGGTTG TCAACTCGTT GTGTTGACCC TTTGATTAAA

2801 GGACACTAGA TGCTAATATC AATGCTGATA TTGGTGAATG GGCTTTTTGG
     CCTGTGATCT ACGATTATAG TTACGACTAT AACCACTTAC CCGAAAAACC

2851 GAAAATAAAA AAATCTCTCC GAACAACTAC GTGGAGAAGA GCTGTCTTTC
     CTTTTATTTT TTTAGAGAGG CTTGTTGATG CACCTCTTCT CGACAGAAAG

2901 GAAACTTTAT CGCTCAACGA GACAGAAGAC GATGATGCGA CATCGTCGAG
     CTTTGAAATA GCGAGTTGCT CTGTCTTCTG CTACTACGCT GTAGCAGCTC

2951 AACTACAAAG GGAAGAATCT CCGACCGGGC CACCAGGAAG TATTCGGACC
     TTGATGTTTC CCTTCTTAGA GGCTGGCCCG GTGGTCCTTC ATAAGCCTGG

3001 TGGTTCCAAA GGATTCCCCT GGGATGGTTT CATTGCACGT ACCAGAAGGG
     ACCAAGGTTT CCTAAGGGGA CCCTACCAAA GTAACGTGCA TGGTCTTCCC

3051 GAAACAACAT TGCCGTCTCA GAATTCGACA GAAGGTCGAA GAGTAGATGT
     CTTTGTTGTA ACGGCAGAGT CTTAAGCTGT CTTCCAGCTT CTCATCTACA

3101 GAATACTCAG GAAACTATCA CAGAGACAAC TGCAACAATC ATAGGCACTA
     CTTATGAGTC CTTTGATAGT GTCTCTGTTG ACGTTGTTAG TATCCGTGAT

3151 ACGGTAACAA CATGCAGATC TCCACCATCG GGACAGGACT GAGCTCCAGC
     TGCCATTGTT GTACGTCTAG AGGTGGTAGC CCTGTCCTGA CTCGAGGTCG

                                NcoI
3201 CAAATCCTGA GTTCCTCACC GACCATGGCA CCAAGCCCTG AGACTCAGAC
     GTTTAGGACT CAAGGAGTGG CTGGTACCGT GGTTCGGGAC TCTGAGTCTG

3251 CTCCACAACC TACACACCAA AACTACCAGT GATGACCACC GAGGAACCAA
     GAGGTGTTGG ATGTGTGGTT TTGATGGTCA CTACTGGTGG CTCCTTGGTT

3301 CAACACCACC GAGAAACTCT CCTGGCTCAA CAACAGAAGC ACCCACTCTC
     GTTGTGGTGG CTCTTTGAGA GGACCGAGTT GTTGTCTTCG TGGGTGAGAG

3351 ACCACCCCAG AGAATATAAC AACAGCGGTT AAAACTGTTT GGGCACAAGA
     TGGTGGGGTC TCTTATATTG TTGTCGCCAA TTTTGACAAA CCCGTGTTCT
```

```
3401  GTCCACAAGC AACGGTCTAA TAACTTCAAC AGTAACAGGT ATTCTTGGGA
      CAGGTGTTCG TTGCCAGATT ATTGAAGTTG TCATTGTCCA TAAGAACCCT
.....................................................................
3451  GCCTTGGACT TCGAAAACGC AGCAGAAGAC AAGTTAACAC CAGGGCCACG
      CGGAACCTGA AGCTTTTGCG TCGTCTTCTG TTCAATTGTG GTCCCGGTGC
.....................................................................
3501  GGTAAATGCA ATCCCAACTT ACACTACTGG ACTGCACAAG AACAACATAA
      CCATTTACGT TAGGGTTGAA TGTGATGACC TGACGTGTTC TTGTTGTATT
.....................................................................
                   BamHI
3551  TGCTGCTGGG ATTGCCTGGA TCCCGTACTT TGGACCGGGT GCAGAAGGCA
      ACGACGACCC TAACGGACCT AGGGCATGAA ACCTGGCCCA CGTCTTCCGT
.....................................................................
3601  TATACACTGA AGGCCTTATG CACAACCAAA ATGCCTTAGT CTGTGGACTC
      ATATGTGACT TCCGGAATAC GTGTTGGTTT TACGGAATCA GACACCTGAG
.....................................................................
3651  AGACAACTTG CAAATGAAAC AACTCAAGCT CTGCAGCTTT TCTTAAGGGC
      TCTGTTGAAC GTTTACTTTG TTGAGTTCGA GACGTCGAAA AGAATTCCCG
.....................................................................
3701  CACGACGGAG CTGCGGACAT ATACCATACT CAATAGGAAG GCCATAGATT
      GTGCTGCCTC GACGCCTGTA TATGGTATGA GTTATCCTTC CGGTATCTAA
.....................................................................
                                       BamHI
3751  TCCTTCTGCG ACGATGGGGC GGGACATGTA GGATCCTGGG ACCAGATTGT
      AGGAAGACGC TGCTACCCCG CCCTGTACAT CCTAGGACCC TGGTCTAACA
.....................................................................
3801  TGCATTGAGC CACATGATTG GACCAAAAAC ATCACTGATA AAATCAACCA
      ACGTAACTCG GTGTACTAAC CTGGTTTTTG TAGTGACTAT TTTAGTTGGT
.....................................................................
3851  AATCATCCAT GATTTCATCG ACAACCCTTT ACCCAATCAG GATAATGATG
      TTAGTAGGTA CTAAAGTAGC TGTTGGGAAA TGGGTTAGTC CTATTACTAC
.....................................................................
                                  BamHI
3901  ATAATTGGTG GACGGGCTGG AGACAGTGGA TCCCTGCAGG AATAGGCATT
      TATTAACCAC CTGCCCGACC TCTGTCACCT AGGGACGTCC TTATCCGTAA
.....................................................................
3951  ACTGGAATTA TTATTGCAAT CATTGCTCTT CTTTGCGTCT GCAAGCTGCT
      TGACCTTAAT AATAACGTTA GTAACGAGAA GAAACGCAGA CGTTCGACGA
.....................................................................
                                                 BamHI
4001  TTGTTGAATA TCAGAATTCC AGCACTGGCG GCCGTTACTA GTGGATCCGA
      AACAACTTAT AGTCTTAAGG TCGTGACCGC CGGCAATGAT CACCTAGGCT
.....................................................................
                             NarI
         BamHI      XbaI     KasI      BamHI
4051  GCTCGGATCC AAGCTCTAGA CCAGGCGCCT GGATCCAGAT CTGCTGTGCC
      CGAGCCTAGG TTCGAGATCT GGTCCGCGGA CCTAGGTCTA GACGACACGG
.....................................................................
4101  TTCTAGTTGC CAGCCATCTG TTGTTTGCCC CTCCCCCGTG CCTTCCTTGA
      AAGATCAACG GTCGGTAGAC AACAAACGGG GAGGGGGCAC GGAAGGAACT
.....................................................................
4151  CCCTGGAAGG TGCCACTCCC ACTGTCCTTT CCTAATAAAA TGAGGAAATT
      GGGACCTTCC ACGGTGAGGG TGACAGGAAA GGATTATTTT ACTCCTTTAA
.....................................................................
```

```
4201  GCATCGCATT GTCTGAGTAG GTGTCATTCT ATTCTGGGGG GTGGGGTGGG
      CGTAGCGTAA CAGACTCATC CACAGTAAGA TAAGACCCCC CACCCCACCC
.......................................................
                                              SphI
4251  GCAGCACAGC AAGGGGGAGG ATTGGGAAGA CAATAGCAGG CATGCTGGGG
      CGTCGTGTCG TTCCCCCTCC TAACCCTTCT GTTATCGTCC GTACGACCCC
.......................................................
                        KpnI
4301  ATGCGGTGGG CTCTATGGGT ACCCAGGTGC TGAAGAATTG ACCCGGTTCC
      TACGCCACCC GAGATACCCA TGGGTCCACG ACTTCTTAAC TGGGCCAAGG
.......................................................
4351  TCCTGGGCCA GAAAGAAGCA GGCACATCCC CTTCTCTGTG ACACACCCTG
      AGGACCCGGT CTTTCTTCGT CCGTGTAGGG GAAGAGACAC TGTGTGGGAC
.......................................................
4401  TCCACGCCCC TGGTTCTTAG TTCCAGCCCC ACTCATAGGA CACTCATAGC
      AGGTGCGGGG ACCAAGAATC AAGGTCGGGG TGAGTATCCT GTGAGTATCG
.......................................................
4451  TCAGGAGGGC TCCGCCTTCA ATCCCACCCG CTAAAGTACT TGGAGCGGTC
      AGTCCTCCCG AGGCGGAAGT TAGGGTGGGC GATTTCATGA ACCTCGCCAG
.......................................................
4501  TCTCCCTCCC TCATCAGCCC ACCAAACCAA ACCTAGCCTC CAAGAGTGGG
      AGAGGGAGGG AGTAGTCGGG TGGTTTGGTT TGGATCGGAG GTTCTCACCC
.......................................................
4551  AAGAAATTAA AGCAAGATAG GCTATTAAGT GCAGAGGGAG AGAAAATGCC
      TTCTTTAATT TCGTTCTATC CGATAATTCA CGTCTCCCTC TCTTTTACGG
.......................................................
                                        XmnI
4601  TCCAACATGT GAGGAAGTAA TGAGAGAAAT CATAGAATTT CTTCCGCTTC
      AGGTTGTACA CTCCTTCATT ACTCTCTTTA GTATCTTAAA GAAGGCGAAG
.......................................................
4651  CTCGCTCACT GACTCGCTGC GCTCGGTCGT TCGGCTGCGG CGAGCGGTAT
      GAGCGAGTGA CTGAGCGACG CGAGCCAGCA AGCCGACGCC GCTCGCCATA
.......................................................
4701  CAGCTCACTC AAAGGCGGTA ATACGGTTAT CCACAGAATC AGGGGATAAC
      GTCGAGTGAG TTTCCGCCAT TATGCCAATA GGTGTCTTAG TCCCCTATTG
.......................................................
4751  GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA GGAACCGTAA
      CGTCCTTTCT TGTACACTCG TTTTCCGGTC GTTTTCCGGT CCTTGGCATT
.......................................................
4801  AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC
      TTTCCGGCGC AACGACCGCA AAAAGGTATC CGAGGCGGGG GGACTGCTCG
.......................................................
4851  ATCACAAAAA TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA
      TAGTGTTTTT AGCTGCGAGT TCAGTCTCCA CCGCTTTGGG CTGTCCTGAT
.......................................................
4901  TAAAGATACC AGGCGTTTCC CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT
      ATTTCTATGG TCCGCAAAGG GGGACCTTCG AGGGAGCACG CGAGAGGACA
.......................................................
4951  TCCGACCCTG CCGCTTACCG GATACCTGTC CGCCTTTCTC CCTTCGGGAA
      AGGCTGGGAC GGCGAATGGC CTATGGACAG GCGGAAAGAG GGAAGCCCTT
.......................................................
5001  GCGTGGCGCT TTCTCAATGC TCACGCTGTA GGTATCTCAG TTCGGTGTAG
      CGCACCGCGA AAGAGTTACG AGTGCGACAT CCATAGAGTC AAGCCACATC
.......................................................
```

```
5051  GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG TTCAGCCCGA
      CAGCAAGCGA GGTTCGACCC GACACACGTG CTTGGGGGC  AAGTCGGGCT

5101  CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC
      GGCGACGCGG AATAGGCCAT TGATAGCAGA ACTCAGGTTG GGCCATTCTG

5151  ACGACTTATC GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG
      TGCTGAATAG CGGTGACCGT CGTCGGTGAC CATTGTCCTA ATCGTCTCGC

5201  AGGTATGTAG GCGGTGCTAC AGAGTTCTTG AAGTGGTGGC CTAACTACGG
      TCCATACATC CGCCACGATG TCTCAAGAAC TTCACCACCG GATTGATGCC

5251  CTACACTAGA AGGACAGTAT TTGGTATCTG CGCTCTGCTG AAGCCAGTTA
      GATGTGATCT TCCTGTCATA AACCATAGAC GCGAGACGAC TTCGGTCAAT

5301  CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT CCGGCAAACA AACCACCGCT
      GGAAGCCTTT TTCTCAACCA TCGAGAACTA GGCCGTTTGT TTGGTGGCGA

5351  GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG CAGATTACGC GCAGAAAAAA
      CCATCGCCAC CAAAAAAACA AACGTTCGTC GTCTAATGCG CGTCTTTTTT

5401  AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT
      TCCTAGAGTT CTTCTAGGAA ACTAGAAAAG ATGCCCCAGA CTGCGAGTCA

5451  GGAACGAAAA CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG
      CCTTGCTTTT GAGTGCAATT CCCTAAAACC AGTACTCTAA TAGTTTTTCC

5501  ATCTTCACCT AGATCCTTTT AAATTAAAAA TGAAGTTTTA AATCAATCTA
      TAGAAGTGGA TCTAGGAAAA TTTAATTTTT ACTTCAAAAT TTAGTTAGAT

5551  AAGTATATAT GAGTAAACTT GGTCTGACAG TTACCAATGC TTAATCAGTG
      TTCATATATA CTCATTTGAA CCAGACTGTC AATGGTTACG AATTAGTCAC

5601  AGGCACCTAT CTCAGCGATC TGTCTATTTC GTTCATCCAT AGTTGCCTGA
      TCCGTGGATA GAGTCGCTAG ACAGATAAAG CAAGTAGGTA TCAACGGACT

5651  CTCCGGGGGG GGGGGGCGCT GAGGTCTGCC TCGTGAAGAA GGTGTTGCTG
      GAGGCCCCCC CCCCCCGCGA CTCCAGACGG AGCACTTCTT CCACAACGAC

5701  ACTCATACCA GGCCTGAATC GCCCCATCAT CCAGCCAGAA AGTGAGGGAG
      TGAGTATGGT CCGGACTTAG CGGGGTAGTA GGTCGGTCTT TCACTCCCTC

5751  CCACGGTTGA TGAGAGCTTT GTTGTAGGTG GACCAGTTGG TGATTTTGAA
      GGTGCCAACT ACTCTCGAAA CAACATCCAC CTGGTCAACC ACTAAAACTT

5801  CTTTTGCTTT GCCACGGAAC GGTCTGCGTT GTCGGGAAGA TGCGTGATCT
      GAAAACGAAA CGGTGCCTTG CCAGACGCAA CAGCCCTTCT ACGCACTAGA

5851  GATCCTTCAA CTCAGCAAAA GTTCGATTTA TTCAACAAAG CCGCCGTCCC
      CTAGGAAGTT GAGTCGTTTT CAAGCTAAAT AAGTTGTTTC GGCGGCAGGG

5901  GTCAAGTCAG CGTAATGCTC TGCCAGTGTT ACAACCAATT AACCAATTCT
      CAGTTCAGTC GCATTACGAG ACGGTCACAA TGTTGGTTAA TTGGTTAAGA

5951  GATTAGAAAA ACTCATCGAG CATCAAATGA AACTGCAATT TATTCATATC
      CTAATCTTTT TGAGTAGCTC GTAGTTTACT TTGACGTTAA ATAAGTATAG
```

```
6001  AGGATTATCA ATACCATATT TTTGAAAAAG CCGTTTCTGT AATGAAGGAG
      TCCTAATAGT TATGGTATAA AAACTTTTTC GGCAAAGACA TTACTTCCTC
...........................................................
6051  AAAACTCACC GAGGCAGTTC CATAGGATGG CAAGATCCTG GTATCGGTCT
      TTTTGAGTGG CTCCGTCAAG GTATCCTACC GTTCTAGGAC CATAGCCAGA
...........................................................
6101  GCGATTCCGA CTCGTCCAAC ATCAATACAA CCTATTAATT TCCCCTCGTC
      CGCTAAGGCT GAGCAGGTTG TAGTTATGTT GGATAATTAA AGGGGAGCAG
...........................................................
6151  AAAAATAAGG TTATCAAGTG AGAAATCACC ATGAGTGACG ACTGAATCCG
      TTTTTATTCC AATAGTTCAC TCTTTAGTGG TACTCACTGC TGACTTAGGC
...........................................................
                 HindIII
6201  GTGAGAATGG CAAAAGCTTA TGCATTTCTT TCCAGACTTG TTCAACAGGC
      CACTCTTACC GTTTTCGAAT ACGTAAAGAA AGGTCTGAAC AAGTTGTCCG
...........................................................
6251  CAGCCATTAC GCTCGTCATC AAAATCACTC GCATCAACCA AACCGTTATT
      GTCGGTAATG CGAGCAGTAG TTTTAGTGAG CGTAGTTGGT TTGGCAATAA
...........................................................
                                        PvuI
6301  CATTCGTGAT TGCGCCTGAC CGAGACGAAA TACGCGATCG CTGTTAAAAG
      GTAAGCACTA ACGCGGACTC GCTCTGCTTT ATGCGCTAGC GACAATTTTC
...........................................................
6351  GACAATTACA AACAGGAATC GAATGCAACC GGCGCAGGAA CACTGCCAGC
      CTGTTAATGT TTGTCCTTAG CTTACGTTGG CCGCGTCCTT GTGACGGTCG
...........................................................
6401  GCATCAACAA TATTTTCACC TGAATCAGGA TATTCTTCTA ATACCTGGAA
      CGTAGTTGTT ATAAAAGTGG ACTTAGTCCT ATAAGAAGAT TATGGACCTT
...........................................................
6451  TGCTGTTTTC CCGGGGATCG CAGTGGTGAG TAACCATGCA TCATCAGGAG
      ACGACAAAAG GGCCCCTAGC GTCACCACTC ATTGGTACGT AGTAGTCCTC
...........................................................
6501  TACGGATAAA ATGCTTGATG GTCGGAAGAG GCATAAATTC CGTCAGCCAG
      ATGCCTATTT TACGAACTAC CAGCCTTCTC CGTATTTAAG GCAGTCGGTC
...........................................................
6551  TTTAGTCTGA CCATCTCATC TGTAACATCA TTGGCAACGC TACCTTTGCC
      AAATCAGACT GGTAGAGTAG ACATTGTAGT AACCGTTGCG ATGGAAACGG
...........................................................
                                                   ClaI
6601  ATGTTTCAGA AACAACTCTG GCGCATCGGG CTTCCCATAC AATCGATAGA
      TACAAAGTCT TTGTTGAGAC CGCGTAGCCC GAAGGGTATG TTAGCTATCT
...........................................................
6651  TTGTCGCACC TGATTGCCCG ACATTATCGC GAGCCCATTT ATACCCATAT
      AACAGCGTGG ACTAACGGGC TGTAATAGCG CTCGGGTAAA TATGGGTATA
...........................................................
6701  AAATCAGCAT CCATGTTGGA ATTTAATCGC GGCCTCGAGC AAGACGTTTC
      TTTAGTCGTA GGTACAACCT TAAATTAGCG CCGGAGCTCG TTCTGCAAAG
...........................................................
6751  CCGTTGAATA TGGCTCATAA CACCCCTTGT ATTACTGTTT ATGTAAGCAG
      GGCAACTTAT ACCGAGTATT GTGGGGAACA TAATGACAAA TACATTCGTC
...........................................................
6801  ACAGTTTTAT TGTTCATGAT GATATATTTT TATCTTGTGC AATGTAACAT
      TGTCAAAATA ACAAGTACTA CTATATAAAA ATAGAACACG TTACATTGTA
...........................................................
```

```
                         DraIII
                        -----------
6851  CAGAGATTTT GAGACACAAC GTGGCTTTCC CCCCCCCCCC ATTATTGAAG
      GTCTCTAAAA CTCTGTGTTG CACCGAAAGG GGGGGGGGGG TAATAACTTC
.......................................................................
6901  CATTTATCAG GGTTATTGTC TCATGAGCGG ATACATATTT GAATGTATTT
      GTAAATAGTC CCAATAACAG AGTACTCGCC TATGTATAAA CTTACATAAA
.......................................................................
6951  AGAAAAATAA ACAAATAGGG GTTCCGCGCA CATTTCCCCG AAAAGTGCCA
      TCTTTTTATT TGTTTATCCC CAAGGCGCGT GTAAAGGGGC TTTTCACGGT
.......................................................................
7001  CCTGACGTCT AAGAAACCAT TATTATCATG ACATTAACCT ATAAAAATAG
      GGACTGCAGA TTCTTTGGTA ATAATAGTAC TGTAATTGGA TATTTTTATC
.......................................................................
7051  GCGTATCACG AGGCCCTTTC GTC
      CGCATAGTGC TCCGGGAAAG CAG
.......................................................................
```

pVR 1012-GP(Z)

Sequence Listing ID No: 3

General Description

DNA pVR 1012-GP(Z)
Local object
Created: 09/15/98 05:06PM
Last Modification Date: ? (no data)
length: 7285 bp
storage type: Basic
form: Circular

Comments

Restriction Map

DraIII: 1 site	CACNNNGTG GTGNNNCAC
HindIII: 1 site	AAGCTT TTCGAA
HpaI: 1 site	GTTAAC CAATTG
KasI: 1 site	GGCGCC CCGCGG
NarI: 1 site	GGCGCC CCGCGG
NotI: 1 site	GCGGCCGC CGCCGGCG
PmlI: 1 site	CACGTG GTGCAC
PvuI: 1 site	CGATCG GCTAGC
SacII: 1 site	CCGCGG GGCGCC
XbaI: 1 site	TCTAGA AGATCT
XhoI: 1 site	CTCGAG GAGCTC
EcoRV: 2 sites	GATATC CTATAG
NcoI: 2 sites	CCATGG GGTACC
NdeI: 2 sites	CATATG GTATAC
SphI: 2 sites	GCATGC CGTACG

Functional Map

CDS (4 signals)

CMV IE 5' UT
Start: 886 End: 1129

CMV IE INT
Start: 1130 End: 1840

TbGH
Start: 4302 End: 4854

Kan r
Start: 6350 End: 6972 (Complementary)

Misc_feature (2 signals)

CMV enhancer

Start: 248 End: 885

GP(Z)

Start: 1870 End: 4301

Annotations

```
  1  TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
     AGCGCGCAAA GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC

 51  GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
     CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT CTGTTCGGGC

101  TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
     AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGACC GAATTGATAC

                                        NdeI
151  CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
     GCCGTAGTCT CGTCTAACAT GACTCTCACG TGGTATACGC CACACTTTAT

201  CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA
     GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCTAACC GATAACCGGT

251  TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
     AACGTATGCA ACATAGGTAT AGTATTATAC ATGTAAATAT AACCGAGTAC

301  TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
     AGGTTGTAAT GGCGGTACAA CTGTAACTAA TAACTGATCA ATAATTATCA

351  AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
     TTAGTTAATG CCCCAGTAAT CAAGTATCGG GTATATACCT CAAGGCGCAA

401  ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG
     TGTATTGAAT GCCATTTACC GGGCGGACCG ACTGGCGGGT TGCTGGGGGC

451  CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
     GGGTAACTGC AGTTATTACT GCATACAAGG GTATCATTGC GGTTATCCCT

501  CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
     GAAAGGTAAC TGCAGTTACC CACCTCATAA ATGCCATTTG ACGGGTGAAC

                            NdeI
551  GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
     CGTCATGTAG TTCACATAGT ATACGGTTCA TGCGGGGGAT AACTGCAGTT

601  TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
     ACTGCCATTT ACCGGGCGGA CCGTAATACG GGTCATGTAC TGGAATACCC

                                                   NcoI
651  ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
     TGAAAGGATG AACCGTCATG TAGATGCATA ATCAGTAGCG ATAATGGTAC

     NcoI
701  GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC
     CACTACGCCA AAACCGTCAT GTAGTTACCC GCACCTATCG CCAAACTGAG

751  ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT
     TGCCCCTAAA GGTTCAGAGG TGGGGTAACT GCAGTTACCC TCAAACAAAA

801  GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA
     CCGTGGTTTT AGTTGCCCTG AAAGGTTTTA CAGCATTGTT GAGGCGGGGT
```

851 TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
AACTGCGTTT ACCCGCCATC CGCACATGCC ACCCTCCAGA TATATTCGTC

901 AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT
TCGAGCAAAT CACTTGGCAG TCTAGCGGAC CTCTGCGGTA GGTGCGACAA

SacII

951 TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA
AACTGGAGGT ATCTTCTGTG GCCCTGGCTA GGTCGGAGGC GCCGGCCCTT

1001 CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC
GCCACGTAAC CTTGCGCCTA AGGGGCACGG TTCTCACTGC ATTCATGGCG

SphI

1051 CTATAGACTC TATAGGCACA CCCCTTTGGC TCTTATGCAT GCTATACTGT
GATATCTGAG ATATCCGTGT GGGGAAACCG AGAATACGTA CGATATGACA

1101 TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA TAGGTGATGG
AAAACCGAAC CCCGGATATG TGGGGGCGAA GGAATACGAT ATCCACTACC

1151 TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
ATATCGAATC GGATATCCAC ACCCAATAAC TGGTAATAAC TGGTGAGGGG

1201 TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA
ATAACCACTG CTATGAAAGG TAATGATTAG GTATTGTACC GAGAAACGGT

1251 CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGTCCTTC AGAGACTGAC
GTTGATAGAG ATAACCGATA TACGGTTATG AGACAGGAAG TCTCTGACTG

1301 ACGGACTCTG TATTTTTACA GGATGGGGTC CCATTTATTA TTTACAAATT
TGCCTGAGAC ATAAAAATGT CCTACCCCAG GGTAAATAAT AAATGTTTAA

1351 CACATATACA ACAACGCCGT CCCCCGTGCC CGCAGTTTTT ATTAAACATA
GTGTATATGT TGTTGCGGCA GGGGGCACGG GCGTCAAAAA TAATTTGTAT

1401 GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA CATGGGCTCT
CGCACCCTAG AGGTGCGCTT AGAGCCCATG CACAAGGCCT GTACCCGAGA

1451 TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
AGAGGCCATC GCCGCCTCGA AGGTGTAGGC TCGGGACCAG GGTACGGAGG

1501 AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG
TCGCCGAGTA CCAGCGAGCC GTCGAGGAAC GAGGATTGTC ACCTCCGGTC

1551 ACTTAGGCAC AGCACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG
TGAATCCGTG TCGTGTTACG GGTGGTGGTG GTCACACGGC GTGTTCCGGC

1601 TGGCGGTAGG GTATGTGTCT GAAAATGAGC GTGGAGATTG GGCTCGCACG
ACCGCCATCC CATACACAGA CTTTTACTCG CACCTCTAAC CCGAGCGTGC

1651 GCTGACGCAG ATGGAAGACT TAAGGCAGCG GCAGAAGAAG ATGCAGGCAG
CGACTGCGTC TACCTTCTGA ATTCCGTCGC CGTCTTCTTC TACGTCCGTC

1701 CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC GTTGCGGTGC
GACTCAACAA CATAAGACTA TTCTCAGTCT CCATTGAGGG CAACGCCACG

```
     HpaI
1751 TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
     ACAATTGCCA CCTCCCGTCA CATCAGACTC GTCATGAGCA ACGACGGCGC
.........................................................................
                                                   NcoI
1801 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT
     GCGCGGTGGT CTGTATTATC GACTGTCTGA TTGTCTGACA AGGAAAGGTA
.........................................................................
     NcoI                               PmlI        EcoRVNotI
1851 GGGTCTTTTC TGCAGTCACC GTCGTCGACA CGTGTGATCA GATATCGCGG
     CCCAGAAAAG ACGTCAGTGG CAGCAGCTGT GCACACTAGT CTATAGCGCC
.........................................................................
                 NarI

     NotI XbaI   KasI
1901 CCGCTCTAGA CCAGGCGCCT GGATCGATCC GCGATGAAGA TTAAGCCGAC
     GGCGAGATCT GGTCCGCGGA CCTAGCTAGG CGCTACTTCT AATTCGGCTG
.........................................................................
1951 AGTGAGCGTA ATCTTCATCT CTCTTAGATT ATTTGTTTTC CAGAGTAGGG
     TCACTCGCAT TAGAAGTAGA GAGAATCTAA TAAACAAAAG GTCTCATCCC
.........................................................................
2001 GTCGTCAGGT CCTTTTCAAT CGTGTAACCA AAATAAACTC CACTAGAAGG
     CAGCAGTCCA GGAAAAGTTA GCACATTGGT TTTATTTGAG GTGATCTTCC
.........................................................................
2051 ATATTGTGGG GCAACAACAC AATGGGCGTT ACAGGAATAT TGCAGTTACC
     TATAACACCC CGTTGTTGTG TTACCCGCAA TGTCCTTATA ACGTCAATGG
.........................................................................
2101 TCGTGATCGA TTCAAGAGGA CATCATTCTT TCTTTGGGTA ATTATCCTTT
     AGCACTAGCT AAGTTCTCCT GTAGTAAGAA AGAAACCCAT TAATAGGAAA
.........................................................................
2151 TCCAAAGAAC ATTTTCCATC CCACTTGGAG TCATCCACAA TAGCACATTA
     AGGTTTCTTG TAAAAGGTAG GGTGAACCTC AGTAGGTGTT ATCGTGTAAT
.........................................................................
2201 CAGGTTAGTG ATGTCGACAA ACTAGTTTGT CGTGACAAAC TGTCATCCAC
     GTCCAATCAC TACAGCTGTT TGATCAAACA GCACTGTTTG ACAGTAGGTG
.........................................................................
2251 AAATCAATTG AGATCAGTTG GACTGAATCT CGAAGGGAAT GGAGTGGCAA
     TTTAGTTAAC TCTAGTCAAC CTGACTTAGA GCTTCCCTTA CCTCACCGTT
.........................................................................
2301 CTGACGTGCC ATCTGCAACT AAAAGATGGG GCTTCAGGTC CGGTGTCCCA
     GACTGCACGG TAGACGTTGA TTTTCTACCC CGAAGTCCAG GCCACAGGGT
.........................................................................
2351 CCAAAGGTGG TCAATTATGA AGCTGGTGAA TGGGCTGAAA ACTGCTACAA
     GGTTTCCACC AGTTAATACT TCGACCACTT ACCCGACTTT TGACGATGTT
.........................................................................
2401 TCTTGAAATC AAAAAACCTG ACGGGAGTGA GTGTCTACCA GCAGCGCCAG
     AGAACTTTAG TTTTTTGGAC TGCCCTCACT CACAGATGGT CGTCGCGGTC
.........................................................................
2451 ACGGGATTCG GGGCTTCCCC CGGTGCCGGT ATGTGCACAA AGTATCAGGA
     TGCCCTAAGC CCCGAAGGGG GCCACGGCCA TACACGTGTT TCATAGTCCT
.........................................................................
2501 ACGGGACCGT GTGCCGGAGA CTTTGCCTTC CATAAAGAGG GTGCTTTCTT
     TGCCCTGGCA CACGGCCTCT GAAACGGAAG GTATTTCTCC CACGAAAGAA
.........................................................................
```

```
2551  CCTGTATGAT CGACTTGCTT CCACAGTTAT CTACCGAGGA ACGACTTTCG
      GGACATACTA GCTGAACGAA GGTGTCAATA GATGGCTCCT TGCTGAAAGC

2601  CTGAAGGTGT CGTTGCATTT CTGATACTGC CCCAAGCTAA GAAGGACTTC
      GACTTCCACA GCAACGTAAA GACTATGACG GGGTTCGATT CTTCCTGAAG

2651  TTCAGCTCAC ACCCCTTGAG AGAGCCGGTC AATGCAACGG AGGACCCGTC
      AAGTCGAGTG TGGGGAACTC TCTCGGCCAG TTACGTTGCC TCCTGGGCAG

                             EcoRV
2701  TAGTGGCTAC TATTCTACCA CAATTAGATA TCAGGCTACC GGTTTTGGAA
      ATCACCGATG ATAAGATGGT GTTAATCTAT AGTCCGATGG CCAAAACCTT

2751  CCAATGAGAC AGAGTACTTG TTCGAGGTTG ACAATTTGAC CTACGTCCAA
      GGTTACTCTG TCTCATGAAC AAGCTCCAAC TGTTAAACTG GATGCAGGTT

2801  CTTGAATCAA GATTCACACC ACAGTTTCTG CTCCAGCTGA ATGAGACAAT
      GAACTTAGTT CTAAGTGTGG TGTCAAAGAC GAGGTCGACT TACTCTGTTA

2851  ATATACAAGT GGGAAAAGGA GCAATACCAC GGGAAAACTA ATTTGGAAGG
      TATATGTTCA CCCTTTTCCT CGTTATGGTG CCCTTTTGAT TAAACCTTCC

2901  TCAACCCCGA AATTGATACA ACAATCGGGG AGTGGGCCTT CTGGGAAACT
      AGTTGGGGCT TTAACTATGT TGTTAGCCCC TCACCCGGAA GACCCTTTGA

2951  AAAAAAAACC TCACTAGAAA AATTCGCAGT GAAGAGTTGT CTTTCACAGT
      TTTTTTTTGG AGTGATCTTT TTAAGCGTCA CTTCTCAACA GAAAGTGTCA

3001  TGTATCAAAC GGAGCCAAAA ACATCAGTGG TCAGAGTCCG GCGCGAACTT
      ACATAGTTTG CCTCGGTTTT TGTAGTCACC AGTCTCAGGC CGCGCTTGAA

3051  CTTCCGACCC AGGGACCAAC ACAACAACTG AAGACCACAA AATCATGGCT
      GAAGGCTGGG TCCCTGGTTG TGTTGTTGAC TTCTGGTGTT TTAGTACCGA

3101  TCAGAAAATT CCTCTGCAAT GGTTCAAGTG CACAGTCAAG GAAGGGAAGC
      AGTCTTTTAA GGAGACGTTA CCAAGTTCAC GTGTCAGTTC CTTCCCTTCG

3151  TGCAGTGTCG CATCTAACAA CCCTTGCCAC AATCTCCACG AGTCCCCAAT
      ACGTCACAGC GTAGATTGTT GGGAACGGTG TTAGAGGTGC TCAGGGGTTA

3201  CCCTCACAAC CAAACCAGGT CCGGACAACA GCACCCATAA TACACCCGTG
      GGGAGTGTTG GTTTGGTCCA GGCCTGTTGT CGTGGGTATT ATGTGGGCAC

3251  TATAAACTTG ACATCTCTGA GGCAACTCAA GTTGAACAAC ATCACCGCAG
      ATATTTGAAC TGTAGAGACT CCGTTGAGTT CAACTTGTTG TAGTGGCGTC

3301  AACAGACAAC GACAGCACAG CCTCCGACAC TCCCTCTGCC ACGACCGCAG
      TTGTCTGTTG CTGTCGTGTC GGAGGCTGTG AGGGAGACGG TGCTGGCGTC

3351  CCGGACCCCC AAAAGCAGAG AACACCAACA CGAGCAAGAG CACTGACTTC
      GGCCTGGGGG TTTTCGTCTC TTGTGGTTGT GCTCGTTCTC GTGACTGAAG

3401  CTGGACCCCG CCACCACAAC AAGTCCCCAA AACCACAGCG AGACCGCTGG
      GACCTGGGGC GGTGGTGTTG TTCAGGGGTT TTGGTGTCGC TCTGGCGACC
```

```
3451  CAACAACAAC ACTCATCACC AAGATACCGG AGAAGAGAGT GCCAGCAGCG
      GTTGTTGTTG TGAGTAGTGG TTCTATGGCC TCTTCTCTCA CGGTCGTCGC
.....................................................................
3501  GGAAGCTAGG CTTAATTACC AATACTATTG CTGGAGTCGC AGGACTGATC
      CCTTCGATCC GAATTAATGG TTATGATAAC GACCTCAGCG TCCTGACTAG
.....................................................................
3551  ACAGGCGGGA GAAGAACTCG AAGAGAAGCA ATTGTCAATG CTCAACCCAA
      TGTCCGCCCT CTTCTTGAGC TTCTCTTCGT TAACAGTTAC GAGTTGGGTT
.....................................................................
3601  ATGCAACCCT AATTTACATT ACTGGACTAC TCAGGATGAA GGTGCTGCAA
      TACGTTGGGA TTAAATGTAA TGACCTGATG AGTCCTACTT CCACGACGTT
.....................................................................
3651  TCGGACTGGC CTGGATACCA TATTTCGGGC CAGCAGCCGA GGGAATTTAC
      AGCCTGACCG GACCTATGGT ATAAAGCCCG GTCGTCGGCT CCCTTAAATG
.....................................................................
3701  ATAGAGGGGC TAATGCACAA TCAAGATGGT TTAATCTGTG GGTTGAGACA
      TATCTCCCCG ATTACGTGTT AGTTCTACCA AATTAGACAC CCAACTCTGT
.....................................................................
3751  GCTGGCCAAC GAGACGACTC AAGCTCTTCA ACTGTTCCTG AGAGCCACAA
      CGACCGGTTG CTCTGCTGAG TTCGAGAAGT TGACAAGGAC TCTCGGTGTT
.....................................................................
3801  CTGAGCTACG CACCTTTTCA ATCCTCAACC GTAAGGCAAT TGATTTCTTG
      GACTCGATGC GTGGAAAAGT TAGGAGTTGG CATTCCGTTA ACTAAAGAAC
.....................................................................
3851  CTGCAGCGAT GGGGCGGCAC ATGCCACATT CTGGGACCGG ACTGCTGTAT
      GACGTCGCTA CCCCGCCGTG TACGGTGTAA GACCCTGGCC TGACGACATA
.....................................................................
3901  CGAACCACAT GATTGGACCA AGAACATAAC AGACAAAATT GATCAGATTA
      GCTTGGTGTA CTAACCTGGT TCTTGTATTG TCTGTTTTAA CTAGTCTAAT
.....................................................................
3951  TTCATGATTT TGTTGATAAA ACCCTTCCGG ACCAGGGGGA CAATGACAAT
      AAGTACTAAA ACAACTATTT TGGGAAGGCC TGGTCCCCCT GTTACTGTTA
.....................................................................
4001  TGGTGGACAG GATGGAGACA ATGGATACCG GCAGGTATTG GAGTTACAGG
      ACCACCTGTC CTACCTCTGT TACCTATGGC CGTCCATAAC CTCAATGTCC
.....................................................................
4051  CGTTATAATT GCAGTTATCG CTTTATTCTG TATATGCAAA TTTGTCTTTT
      GCAATATTAA CGTCAATAGC GAAATAAGAC ATATACGTTT AAACAGAAAA
.....................................................................
4101  AGTTTTTCTT CAGATTGCTT CATGGAAAAG CTCAGCCTCA AATCAATGAA
      TCAAAAAGAA GTCTAACGAA GTACCTTTTC GAGTCGGAGT TTAGTTACTT
.....................................................................
4151  ACCAGGATTT AATTATATGG ATTACTTGAA TCTAAGATTA CTTGACAAAT
      TGGTCCTAAA TTAATATACC TAATGAACTT AGATTCTAAT GAACTGTTTA
.....................................................................
4201  GATAATATAA TACACTGGAG CTTTAAACAT AGCCAATGTG ATTCTAACTC
      CTATTATATT ATGTGACCTC GAAATTTGTA TCGGTTACAC TAAGATTGAG
.....................................................................
4251  CTTTAAACTC ACAGTTAATC ATAAACAAGG TTTGGTACCG AGCTCGAATT
      GAAATTTGAG TGTCAATTAG TATTTGTTCC AAACCATGGC TCGAGCTTAA
.....................................................................
4301  ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTTGTTTGC CCCTCCCCCG
      TAGACGACAC GGAAGATCAA CGGTCGGTAG ACAACAAACG GGGAGGGGGC
.....................................................................
4351  TGCCTTCCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT TTCCTAATAA
      ACGGAAGGAA CTGGGACCTT CCACGGTGAG GGTGACAGGA AAGGATTATT
.....................................................................
```

```
4401  AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT CTATTCTGGG
      TTACTCCTTT AACGTAGCGT AACAGACTCA TCCACAGTAA GATAAGACCC
      ..................................................

4451  GGGTGGGGTG GGGCAGCACA GCAAGGGGGA GGATTGGGAA GACAATAGCA
      CCCACCCCAC CCCGTCGTGT CGTTCCCCCT CCTAACCCTT CTGTTATCGT
      ..................................................

       SphI
      ______
4501  GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCCAGGT GCTGAAGAAT
      CCGTACGACC CCTACGCCAC CCGAGATACC CATGGGTCCA CGACTTCTTA
      ..................................................

4551  TGACCCGGTT CCTCCTGGGC CAGAAAGAAG CAGGCACATC CCCTTCTCTG
      ACTGGGCCAA GGAGGACCCG GTCTTTCTTC GTCCGTGTAG GGGAAGAGAC
      ..................................................

4601  TGACACACCC TGTCCACGCC CCTGGTTCTT AGTTCCAGCC CCACTCATAG
      ACTGTGTGGG ACAGGTGCGG GGACCAAGAA TCAAGGTCGG GGTGAGTATC
      ..................................................

4651  GACACTCATA GCTCAGGAGG GCTCCGCCTT CAATCCCACC CGCTAAAGTA
      CTGTGAGTAT CGAGTCCTCC CGAGGCGGAA GTTAGGGTGG GCGATTTCAT
      ..................................................

4701  CTTGGAGCGG TCTCTCCCTC CCTCATCAGC CCACCAAACC AAACCTAGCC
      GAACCTCGCC AGAGAGGGAG GGAGTAGTCG GGTGGTTTGG TTTGGATCGG
      ..................................................

4751  TCCAAGAGTG GGAAGAAATT AAAGCAAGAT AGGCTATTAA GTGCAGAGGG
      AGGTTCTCAC CCTTCTTTAA TTTCGTTCTA TCCGATAATT CACGTCTCCC
      ..................................................

4801  AGAGAAAATG CCTCCAACAT GTGAGGAAGT AATGAGAGAA ATCATAGAAT
      TCTCTTTTAC GGAGGTTGTA CACTCCTTCA TTACTCTCTT TAGTATCTTA
      ..................................................

4851  TTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC
      AAGAAGGCGA AGGAGCGAGT GACTGAGCGA CGCGAGCCAG CAAGCCGACG
      ..................................................

4901  GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA
      CCGCTCGCCA TAGTCGAGTG AGTTTCCGCC ATTATGCCAA TAGGTGTCTT
      ..................................................

4951  TCAGGGGATA ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC
      AGTCCCCTAT TGCGTCCTTT CTTGTACACT CGTTTTCCGG TCGTTTTCCG
      ..................................................

5001  CAGGAACCGT AAAAAGGCCG CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC
      GTCCTTGGCA TTTTTCCGGC GCAACGACCG CAAAAAGGTA TCCGAGGCGG
      ..................................................

5051  CCCCTGACGA GCATCACAAA AATCGACGCT CAAGTCAGAG GTGGCGAAAC
      GGGGACTGCT CGTAGTGTTT TTAGCTGCGA GTTCAGTCTC CACCGCTTTG
      ..................................................

5101  CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA GCTCCCTCGT
      GGCTGTCCTG ATATTTCTAT GGTCCGCAAA GGGGGACCTT CGAGGGAGCA
      ..................................................

5151  GCGCTCTCCT GTTCCGACCC TGCCGCTTAC CGGATACCTG TCCGCCTTTC
      CGCGAGAGGA CAAGGCTGGG ACGGCGAATG GCCTATGGAC AGGCGGAAAG
      ..................................................

5201  TCCCTTCGGG AAGCGTGGCG CTTTCTCAAT GCTCACGCTG TAGGTATCTC
      AGGGAAGCCC TTCGCACCGC GAAAGAGTTA CGAGTGCGAC ATCCATAGAG
      ..................................................

5251  AGTTCGGTGT AGGTCGTTCG CTCCAAGCTG GGCTGTGTGC ACGAACCCCC
      TCAAGCCACA TCCAGCAAGC GAGGTTCGAC CCGACACACG TGCTTGGGGG
      ..................................................
```

5301 CGTTCAGCCC GACCGCTGCG CCTTATCCGG TAACTATCGT CTTGAGTCCA
GCAAGTCGGG CTGGCGACGC GGAATAGGCC ATTGATAGCA GAACTCAGGT

5351 ACCCGGTAAG ACACGACTTA TCGCCACTGG CAGCAGCCAC TGGTAACAGG
TGGGCCATTC TGTGCTGAAT AGCGGTGACC GTCGTCGGTG ACCATTGTCC

5401 ATTAGCAGAG CGAGGTATGT AGGCGGTGCT ACAGAGTTCT TGAAGTGGTG
TAATCGTCTC GCTCCATACA TCCGCCACGA TGTCTCAAGA ACTTCACCAC

5451 GCCTAACTAC GGCTACACTA GAAGGACAGT ATTTGGTATC TGCGCTCTGC
CGGATTGATG CCGATGTGAT CTTCCTGTCA TAAACCATAG ACGCGAGACG

5501 TGAAGCCAGT TACCTTCGGA AAAAGAGTTG GTAGCTCTTG ATCCGGCAAA
ACTTCGGTCA ATGGAAGCCT TTTTCTCAAC CATCGAGAAC TAGGCCGTTT

5551 CAAACCACCG CTGGTAGCGG TGGTTTTTTT GTTTGCAAGC AGCAGATTAC
GTTTGGTGGC GACCATCGCC ACCAAAAAAA CAAACGTTCG TCGTCTAATG

5601 GCGCAGAAAA AAAGGATCTC AAGAAGATCC TTTGATCTTT TCTACGGGGT
CGCGTCTTTT TTTCCTAGAG TTCTTCTAGG AAACTAGAAA AGATGCCCCA

5651 CTGACGCTCA GTGGAACGAA AACTCACGTT AAGGGATTTT GGTCATGAGA
GACTGCGAGT CACCTTGCTT TTGAGTGCAA TTCCCTAAAA CCAGTACTCT

5701 TTATCAAAAA GGATCTTCAC CTAGATCCTT TTAAATTAAA AATGAAGTTT
AATAGTTTTT CCTAGAAGTG GATCTAGGAA AATTTAATTT TTACTTCAAA

5751 TAAATCAATC TAAAGTATAT ATGAGTAAAC TTGGTCTGAC AGTTACCAAT
ATTTAGTTAG ATTTCATATA TACTCATTTG AACCAGACTG TCAATGGTTA

5801 GCTTAATCAG TGAGGCACCT ATCTCAGCGA TCTGTCTATT TCGTTCATCC
CGAATTAGTC ACTCCGTGGA TAGAGTCGCT AGACAGATAA AGCAAGTAGG

5851 ATAGTTGCCT GACTCCGGGG GGGGGGGGCG CTGAGGTCTG CCTCGTGAAG
TATCAACGGA CTGAGGCCCC CCCCCCCCGC GACTCCAGAC GGAGCACTTC

5901 AAGGTGTTGC TGACTCATAC CAGGCCTGAA TCGCCCCATC ATCCAGCCAG
TTCCACAACG ACTGAGTATG GTCCGGACTT AGCGGGGTAG TAGGTCGGTC

5951 AAAGTGAGGG AGCCACGGTT GATGAGAGCT TTGTTGTAGG TGGACCAGTT
TTTCACTCCC TCGGTGCCAA CTACTCTCGA AACAACATCC ACCTGGTCAA

6001 GGTGATTTTG AACTTTTGCT TTGCCACGGA ACGGTCTGCG TTGTCGGGAA
CCACTAAAAC TTGAAAACGA AACGGTGCCT TGCCAGACGC AACAGCCCTT

6051 GATGCGTGAT CTGATCCTTC AACTCAGCAA AAGTTCGATT TATTCAACAA
CTACGCACTA GACTAGGAAG TTGAGTCGTT TTCAAGCTAA ATAAGTTGTT

6101 AGCCGCCGTC CCGTCAAGTC AGCGTAATGC TCTGCCAGTG TTACAACCAA
TCGGCGGCAG GGCAGTTCAG TCGCATTACG AGACGGTCAC AATGTTGGTT

6151 TTAACCAATT CTGATTAGAA AAACTCATCG AGCATCAAAT GAAACTGCAA
AATTGGTTAA GACTAATCTT TTTGAGTAGC TCGTAGTTTA CTTTGACGTT

6201 TTTATTCATA TCAGGATTAT CAATACCATA TTTTTGAAAA AGCCGTTTCT
AAATAAGTAT AGTCCTAATA GTTATGGTAT AAAAACTTTT TCGGCAAAGA

```
6251  GTAATGAAGG AGAAAACTCA CCGAGGCAGT TCCATAGGAT GGCAAGATCC
      CATTACTTCC TCTTTTGAGT GGCTCCGTCA AGGTATCCTA CCGTTCTAGG

6301  TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC AACCTATTAA
      ACCATAGCCA GACGCTAAGG CTGAGCAGGT TGTAGTTATG TTGGATAATT

6351  TTTCCCCTCG TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA
      AAAGGGGAGC AGTTTTTATT CCAATAGTTC ACTCTTTAGT GGTACTCACT

                                HindIII
6401  CGACTGAATC CGGTGAGAAT GGCAAAAGCT TATGCATTTC TTTCCAGACT
      GCTGACTTAG GCCACTCTTA CCGTTTTCGA ATACGTAAAG AAAGGTCTGA

6451  TGTTCAACAG GCCAGCCATT ACGCTCGTCA TCAAAATCAC TCGCATCAAC
      ACAAGTTGTC CGGTCGGTAA TGCGAGCAGT AGTTTTAGTG AGCGTAGTTG

                                                     PvuI
6501  CAAACCGTTA TTCATTCGTG ATTGCGCCTG AGCGAGACGA AATACGCGAT
      GTTTGGCAAT AAGTAAGCAC TAACGCGGAC TCGCTCTGCT TTATGCGCTA

      PvuI
6551  CGCTGTTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA CCGGCGCAGG
      GCGACAATTT TCCTGTTAAT GTTTGTCCTT AGCTTACGTT GGCCGCGTCC

6601  AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG GATATTCTTC
      TTGTGACGGT CGCGTAGTTG TTATAAAAGT GGACTTAGTC CTATAAGAAG

6651  TAATACCTGG AATGCTGTTT TCCCGGGGAT CGCAGTGGTG AGTAACCATG
      ATTATGGACC TTACGACAAA AGGGCCCCTA GCGTCACCAC TCATTGGTAC

6701  CATCATCAGG AGTACGGATA AAATGCTTGA TGGTCGGAAG AGGCATAAAT
      GTAGTAGTCC TCATGCCTAT TTTACGAACT ACCAGCCTTC TCCGTATTTA

6751  TCCGTCAGCC AGTTTAGTCT GACCATCTCA TCTGTAACAT CATTGGCAAC
      AGGCAGTCGG TCAAATCAGA CTGGTAGAGT AGACATTGTA GTAACCGTTG

6801  GCTACCTTTG CCATGTTTCA GAAACAACTC TGGCGCATCG GGCTTCCCAT
      CGATGGAAAC GGTACAAAGT CTTTGTTGAG ACCGCGTAGC CCGAAGGGTA

6851  ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC GCGAGCCCAT
      TGTTAGCTAT CTAACAGCGT GGACTAACGG GCTGTAATAG CGCTCGGGTA

                                                     XhoI
6901  TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC GCGGCCTCGA
      AATATGGGTA TATTTAGTCG TAGGTACAAC CTTAAATTAG CGCCGGAGCT

      XhoI
6951  GCAAGACGTT TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT
      CGTTCTGCAA AGGGCAACTT ATACCGAGTA TTGTGGGGAA CATAATGACA

7001  TTATGTAAGC AGACAGTTTT ATTGTTCATG ATGATATATT TTTATCTTGT
      AATACATTCG TCTGTCAAAA TAACAAGTAC TACTATATAA AAATAGAACA
```

```
                                      DraIII
7051  GCAATGTAAC ATCAGAGATT TTGAGACACA ACGTGGCTTT CCCCCCCCCC
      CGTTACATTG TAGTCTCTAA AACTCTGTGT TGCACCGAAA GGGGGGGGGG
..............................................................
7101  CCATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC GGATACATAT
      GGTAATAACT TCGTAAATAG TCCCAATAAC AGAGTACTCG CCTATGTATA
..............................................................
7151  TTGAATGTAT TTAGAAAAAT AAACAAATAG GGGTTCCGCG CACATTTCCC
      AACTTACATA AATCTTTTTA TTTGTTTATC CCCAAGGCGC GTGTAAAGGG
..............................................................
7201  CGAAAAGTGC CACCTGACGT CTAAGAAACC ATTATTATCA TGACATTAAC
      GCTTTTCACG GTGGACTGCA GATTCTTTGG TAATAATAGT ACTGTAATTG
..............................................................
7251  CTATAAAAAT AGGCGTATCA CGAGGCCCTT TCGTC
      GATATTTTTA TCCGCATAGT GCTCCGGGAA AGCAG
..............................................................
```

pVR 1012-SGP(Z)

Sequence Listing ID No: 4

General Description

DNA pVR 1012-SGP(Z)
Local object
Created: 09/14/98 04:29PM
Last Modified: 09/15/98 04:50PM
length: 7272 bp
storage type: Basic
form: Circular

Comments

Restriction Map

DraIII: 1 site	CACNNNGTG GTGNNNCAC
HindIII: 1 site	AAGCTT TTCGAA
HpaI: 1 site	GTTAAC CAATTG
KpnI: 1 site	GGTACC CCATGG
NotI: 1 site	GCGGCCGC CGCCGGCG
PmlI: 1 site	CACGTG GTGCAC
PvuI: 1 site	CGATCG GCTAGC
SacII: 1 site	CCGCGG GGCGCC
XbaI: 1 site	TCTAGA AGATCT
XhoI: 1 site	CTCGAG GAGCTC
EcoRV: 2 sites	GATATC CTATAG
NcoI: 2 sites	CCATGG GGTACC
NdeI: 2 sites	CATATG GTATAC
SphI: 2 sites	GCATGC CGTACG

Functional Map

CDS (4 signals)

CMV IE 5' UT
Start: 886 End: 1129

CMV IE INT
Start: 1130 End: 1840

TbGH
Start: 4289 End: 4841

Kan r
Start: 6337 End: 6959 (Complementary)

Misc_feature (2 signals)

CMV enhancer
Start: 248 End: 885
SGP(Z)
Start: 1870 End: 4288

Annotations

```
  1  TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG
     AGCGCGCAAA GCCACTACTG CCACTTTTGG AGACTGTGTA CGTCGAGGGC

 51  GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG
     CTCTGCCAGT GTCGAACAGA CATTCGCCTA CGGCCCTCGT CTGTTCGGGC

101  TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCTGG CTTAACTATG
     AGTCCCGCGC AGTCGCCCAC AACCGCCCAC AGCCCCGACC GAATTGATAC

                                        NdeI
151  CGGCATCAGA GCAGATTGTA CTGAGAGTGC ACCATATGCG GTGTGAAATA
     GCCGTAGTCT CGTCTAACAT GACTCTCACG TGGTATACGC CACACTTTAT

201  CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGATTGG CTATTGGCCA
     GGCGTGTCTA CGCATTCCTC TTTTATGGCG TAGTCTAACC GATAACCGGT

251  TTGCATACGT TGTATCCATA TCATAATATG TACATTTATA TTGGCTCATG
     AACGTATGCA ACATAGGTAT AGTATTATAC ATGTAAATAT AACCGAGTAC

301  TCCAACATTA CCGCCATGTT GACATTGATT ATTGACTAGT TATTAATAGT
     AGGTTGTAAT GGCGGTACAA CTGTAACTAA TAACTGATCA ATAATTATCA

351  AATCAATTAC GGGGTCATTA GTTCATAGCC CATATATGGA GTTCCGCGTT
     TTAGTTAATG CCCCAGTAAT CAAGTATCGG GTATATACCT CAAGGCGCAA

401  ACATAACTTA CGGTAAATGG CCCGCCTGGC TGACCGCCCA ACGACCCCCG
     TGTATTGAAT GCCATTTACC GGGCGGACCG ACTGGCGGGT TGCTGGGGGC

451  CCCATTGACG TCAATAATGA CGTATGTTCC CATAGTAACG CCAATAGGGA
     GGGTAACTGC AGTTATTACT GCATACAAGG GTATCATTGC GGTTATCCCT

501  CTTTCCATTG ACGTCAATGG GTGGAGTATT TACGGTAAAC TGCCCACTTG
     GAAAGGTAAC TGCAGTTACC CACCTCATAA ATGCCATTTG ACGGGTGAAC

                        NdeI
551  GCAGTACATC AAGTGTATCA TATGCCAAGT ACGCCCCCTA TTGACGTCAA
     CGTCATGTAG TTCACATAGT ATACGGTTCA TGCGGGGGAT AACTGCAGTT

601  TGACGGTAAA TGGCCCGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
     ACTGCCATTT ACCGGGCGGA CCGTAATACG GGTCATGTAC TGGAATACCC

                                                    NcoI
651  ACTTTCCTAC TTGGCAGTAC ATCTACGTAT TAGTCATCGC TATTACCATG
     TGAAAGGATG AACCGTCATG TAGATGCATA ATCAGTAGCG ATAATGGTAC

     NcoI
701  GTGATGCGGT TTTGGCAGTA CATCAATGGG CGTGGATAGC GGTTTGACTC
     CACTACGCCA AAACCGTCAT GTAGTTACCC GCACCTATCG CCAAACTGAG

751  ACGGGGATTT CCAAGTCTCC ACCCCATTGA CGTCAATGGG AGTTTGTTTT
     TGCCCCTAAA GGTTCAGAGG TGGGGTAACT GCAGTTACCC TCAAACAAAA

801  GGCACCAAAA TCAACGGGAC TTTCCAAAAT GTCGTAACAA CTCCGCCCCA
     CCGTGGTTTT AGTTGCCCTG AAAGGTTTTA CAGCATTGTT GAGGCGGGGT
```

```
 851  TTGACGCAAA TGGGCGGTAG GCGTGTACGG TGGGAGGTCT ATATAAGCAG
      AACTGCGTTT ACCCGCCATC CGCACATGCC ACCCTCCAGA TATATTCGTC

 901  AGCTCGTTTA GTGAACCGTC AGATCGCCTG GAGACGCCAT CCACGCTGTT
      TCGAGCAAAT CACTTGGCAG TCTAGCGGAC CTCTGCGGTA GGTGCGACAA

                                          SacII
 951  TTGACCTCCA TAGAAGACAC CGGGACCGAT CCAGCCTCCG CGGCCGGGAA
      AACTGGAGGT ATCTTCTGTG GCCCTGGCTA GGTCGGAGGC GCCGGCCCTT

1001  CGGTGCATTG GAACGCGGAT TCCCCGTGCC AAGAGTGACG TAAGTACCGC
      GCCACGTAAC CTTGCGCCTA AGGGGCACGG TTCTCACTGC ATTCATGGCG

                                          SphI
1051  CTATAGACTC TATAGGCACA CCCCTTTGGC TCTTATGCAT GCTATACTGT
      GATATCTGAG ATATCCGTGT GGGGAAACCG AGAATACGTA CGATATGACA

1101  TTTTGGCTTG GGGCCTATAC ACCCCCGCTT CCTTATGCTA TAGGTGATGG
      AAAACCGAAC CCCGGATATG TGGGGGCGAA GGAATACGAT ATCCACTACC

1151  TATAGCTTAG CCTATAGGTG TGGGTTATTG ACCATTATTG ACCACTCCCC
      ATATCGAATC GGATATCCAC ACCCAATAAC TGGTAATAAC TGGTGAGGGG

1201  TATTGGTGAC GATACTTTCC ATTACTAATC CATAACATGG CTCTTTGCCA
      ATAACCACTG CTATGAAAGG TAATGATTAG GTATTGTACC GAGAAACGGT

1251  CAACTATCTC TATTGGCTAT ATGCCAATAC TCTGTCCTTC AGAGACTGAC
      GTTGATAGAG ATAACCGATA TACGGTTATG AGACAGGAAG TCTCTGACTG

1301  ACGGACTCTG TATTTTTACA GGATGGGGTC CCATTTATTA TTTACAAATT
      TGCCTGAGAC ATAAAAATGT CCTACCCCAG GGTAAATAAT AAATGTTTAA

1351  CACATATACA ACAACGCCGT CCCCCGTGCC CGCAGTTTTT ATTAAACATA
      GTGTATATGT TGTTGCGGCA GGGGGCACGG GCGTCAAAAA TAATTTGTAT

1401  GCGTGGGATC TCCACGCGAA TCTCGGGTAC GTGTTCCGGA CATGGGCTCT
      CGCACCCTAG AGGTGCGCTT AGAGCCCATG CACAAGGCCT GTACCCGAGA

1451  TCTCCGGTAG CGGCGGAGCT TCCACATCCG AGCCCTGGTC CCATGCCTCC
      AGAGGCCATC GCCGCCTCGA AGGTGTAGGC TCGGGACCAG GGTACGGAGG

1501  AGCGGCTCAT GGTCGCTCGG CAGCTCCTTG CTCCTAACAG TGGAGGCCAG
      TCGCCGAGTA CCAGCGAGCC GTCGAGGAAC GAGGATTGTC ACCTCCGGTC

1551  ACTTAGGCAC AGCACAATGC CCACCACCAC CAGTGTGCCG CACAAGGCCG
      TGAATCCGTG TCGTGTTACG GGTGGTGGTG GTCACACGGC GTGTTCCGGC

1601  TGGCGGTAGG GTATGTGTCT GAAAATGAGC GTGGAGATTG GGCTCGCACG
      ACCGCCATCC CATACACAGA CTTTTACTCG CACCTCTAAC CCGAGCGTGC

1651  GCTGACGCAG ATGGAAGACT TAAGGCAGCG GCAGAAGAAG ATGCAGGCAG
      CGACTGCGTC TACCTTCTGA ATTCCGTCGC CGTCTTCTTC TACGTCCGTC

1701  CTGAGTTGTT GTATTCTGAT AAGAGTCAGA GGTAACTCCC GTTGCGGTGC
      GACTCAACAA CATAAGACTA TTCTCAGTCT CCATTGAGGG CAACGCCACG
```

```
     HpaI
1751 TGTTAACGGT GGAGGGCAGT GTAGTCTGAG CAGTACTCGT TGCTGCCGCG
     ACAATTGCCA CCTCCCGTCA CATCAGACTC GTCATGAGCA ACGACGGCGC
.....................................................................
                                                     NcoI
1801 CGCGCCACCA GACATAATAG CTGACAGACT AACAGACTGT TCCTTTCCAT
     GCGCGGTGGT CTGTATTATC GACTGTCTGA TTGTCTGACA AGGAAAGGTA
.....................................................................
     NcoI                          PmlI       EcoRVNotI
1851 GGGTCTTTTC TGCAGTCACC GTCGTCGACA CGTGTGATCA GATATCGCGG
     CCCAGAAAAG ACGTCAGTGG CAGCAGCTGT GCACACTAGT CTATAGCGCC
.....................................................................
     NotI XbaI
1901 CCGCTCTAGA CCAGGCGCCT GGATCGAATT GATGAAGATT AAGCCGACAG
     GGCGAGATCT GGTCCGCGGA CCTAGCTTAA CTACTTCTAA TTCGGCTGTC
.....................................................................
1951 TGAGCGTAAT CTTCATCTCT CTTAGATTAT TTGTTTTCCA GAGTAGGGGT
     ACTCGCATTA GAAGTAGAGA GAATCTAATA AACAAAAGGT CTCATCCCCA
.....................................................................
2001 CGTCAGGTCC TTTTCAATCG TGTAACCAAA ATAAACTCCA CTAGAAGGAT
     GCAGTCCAGG AAAAGTTAGC ACATTGGTTT TATTTGAGGT GATCTTCCTA
.....................................................................
2051 ATTGTGGGGC AACAACACAA TGGGCGTTAC AGGAATATTG CAGTTACCTC
     TAACACCCCG TTGTTGTGTT ACCCGCAATG TCCTTATAAC GTCAATGGAG
.....................................................................
2101 GTGATCGATT CAAGAGGACA TCATTCTTTC TTTGGGTAAT TATCCTTTTC
     CACTAGCTAA GTTCTCCTGT AGTAAGAAAG AAACCCATTA ATAGGAAAAG
.....................................................................
2151 CAAAGAACAT TTTCCATCCC ACTTGGAGTC ATCCACAATA GCACATTACA
     GTTTCTTGTA AAAGGTAGGG TGAACCTCAG TAGGTGTTAT CGTGTAATGT
.....................................................................
2201 GGTTAGTGAT GTCGACAAAC TAGTTTGTCG TGACAAACTG TCATCCACAA
     CCAATCACTA CAGCTGTTTG ATCAAACAGC ACTGTTTGAC AGTAGGTGTT
.....................................................................
2251 ATCAATTGAG ATCAGTTGGA CTGAATCTCG AAGGGAATGG AGTGGCAACT
     TAGTTAACTC TAGTCAACCT GACTTAGAGC TTCCCTTACC TCACCGTTGA
.....................................................................
2301 GACGTGCCAT CTGCAACTAA AAGATGGGGC TTCAGGTCCG GTGTCCCACC
     CTGCACGGTA GACGTTGATT TTCTACCCCG AAGTCCAGGC CACAGGGTGG
.....................................................................
2351 AAAGGTGGTC AATTATGAAG CTGGTGAATG GGCTGAAAAC TGCTACAATC
     TTTCCACCAG TTAATACTTC GACCACTTAC CCGACTTTTG ACGATGTTAG
.....................................................................
2401 TTGAAATCAA AAAACCTGAC GGGAGTGAGT GTCTACCAGC AGCGCCAGAC
     AACTTTAGTT TTTTGGACTG CCCTCACTCA CAGATGGTCG TCGCGGTCTG
.....................................................................
2451 GGGATTCGGG GCTTCCCCCG GTGCCGGTAT GTGCACAAAG TATCAGGAAC
     CCCTAAGCCC CGAAGGGGGC CACGGCCATA CACGTGTTTC ATAGTCCTTG
.....................................................................
2501 GGGACCGTGT GCCGGAGACT TTGCCTTCCA TAAAGAGGGT GCTTTCTTCC
     CCCTGGCACA CGGCCTCTGA AACGGAAGGT ATTTCTCCCA CGAAAGAAGG
.....................................................................
2551 TGTATGATCG ACTTGCTTCC ACAGTTATCT ACCGAGGAAC GACTTTCGCT
     ACATACTAGC TGAACGAAGG TGTCAATAGA TGGCTCCTTG CTGAAAGCGA
.....................................................................
```

```
2601  GAAGGTGTCG TTGCATTTCT GATACTGCCC CAAGCTAAGA AGGACTTCTT
      CTTCCACAGC AACGTAAAGA CTATGACGGG GTTCGATTCT TCCTGAAGAA

2651  CAGCTCACAC CCCTTGAGAG AGCCGGTCAA TGCAACGGAG GACCCGTCTA
      GTCGAGTGTG GGGAACTCTC TCGGCCAGTT ACGTTGCCTC CTGGGCAGAT

                                  EcoRV
                                 -------
2701  GTGGCTACTA TTCTACCACA ATTAGATATC AGGCTACCGG TTTTGGAACC
      CACCGATGAT AAGATGGTGT TAATCTATAG TCCGATGGCC AAAACCTTGG

2751  AATGAGACAG AGTACTTGTT CGAGGTTGAC AATTTGACCT ACGTCCAACT
      TTACTCTGTC TCATGAACAA GCTCCAACTG TTAAACTGGA TGCAGGTTGA

2801  TGAATCAAGA TTCACACCAC AGTTTCTGCT CCAGCTGAAT GAGACAATAT
      ACTTAGTTCT AAGTGTGGTG TCAAAGACGA GGTCGACTTA CTCTGTTATA

2851  ATACAAGTGG GAAAAGGAGC AATACCACGG GAAAACTAAT TTGGAAGGTC
      TATGTTCACC CTTTTCCTCG TTATGGTGCC CTTTTGATTA AACCTTCCAG

2901  AACCCCGAAA TTGATACAAC AATCGGGGAG TGGGCCTTCT GGGAAACTAA
      TTGGGGCTTT AACTATGTTG TTAGCCCCTC ACCCGGAAGA CCCTTTGATT

2951  AAAAACCTCA CTAGAAAAAT TCGCAGTGAA GAGTTGTCTT TCACAGTTGT
      TTTTTGGAGT GATCTTTTTA AGCGTCACTT CTCAACAGAA AGTGTCAACA

3001  ATCAAACGGA GCCAAAAACA TCAGTGGTCA GAGTCCGGCG CGAACTTCTT
      TAGTTTGCCT CGGTTTTTGT AGTCACCAGT CTCAGGCCGC GCTTGAAGAA

3051  CCGACCCAGG GACCAACACA ACAACTGAAG ACCACAAAAT CATGGCTTCA
      GGCTGGGTCC CTGGTTGTGT TGTTGACTTC TGGTGTTTTA GTACCGAAGT

3101  GAAAATTCCT CTGCAATGGT TCAAGTGCAC AGTCAAGGAA GGGAAGCTGC
      CTTTTAAGGA GACGTTACCA AGTTCACGTG TCAGTTCCTT CCCTTCGACG

3151  AGTGTCGCAT CTAACAACCC TTGCCACAAT CTCCACGAGT CCCCAATCCC
      TCACAGCGTA GATTGTTGGG AACGGTGTTA GAGGTGCTCA GGGGTTAGGG

3201  TCACAACCAA ACCAGGTCCG GACAACAGCA CCCATAATAC ACCCGTGTAT
      AGTGTTGGTT TGGTCCAGGC CTGTTGTCGT GGGTATTATG TGGGCACATA

3251  AAACTTGACA TCTCTGAGGC AACTCAAGTT GAACAACATC ACCGCAGAAC
      TTTGAACTGT AGAGACTCCG TTGAGTTCAA CTTGTTGTAG TGGCGTCTTG

3301  AGACAACGAC AGCACAGCCT CCGACACTCC CTCTGCCACG ACCGCAGCCG
      TCTGTTGCTG TCGTGTCGGA GGCTGTGAGG GAGACGGTGC TGGCGTCGGC

3351  GACCCCCAAA AGCAGAGAAC ACCAACACGA GCAAGAGCAC TGACTTCCTC
      CTGGGGGTTT TCGTCTCTTG TGGTTGTGCT CGTTCTCGTG ACTGAAGGAC

3401  GACCCCGCCA CCACAACAAG TCCCCAAAAC CACAGCGAGA CCGCTGGCAA
      CTGGGGCGGT GGTGTTGTTC AGGGGTTTTG GTGTCGCTCT GGCGACCGTT

3451  CAACAACACT CATCACCAAG ATACCGGAGA AGAGAGTGCC AGCAGCGGGA
      GTTGTTGTGA GTAGTGGTTC TATGGCCTCT TCTCTCACGG TCGTCGCCCT
```

```
3501  AGCTAGGCTT AATTACCAAT ACTATTGCTG GAGTCGCAGG ACTGATCACA
      TCGATCCGAA TTAATGGTTA TGATAACGAC CTCAGCGTCC TGACTAGTGT

3551  GGCGGGAGAA GAACTCGAAG AGAAGCAATT GTCAATGCTC AACCCAAATG
      CCGCCCTCTT CTTGAGCTTC TCTTCGTTAA CAGTTACGAG TTGGGTTTAC

3601  CAACCCTAAT TTACATTACT GGACTACTCA GGATGAAGGT GCTGCAATCG
      GTTGGGATTA AATGTAATGA CCTGATGAGT CCTACTTCCA CGACGTTAGC

3651  GACTGGCCTG GATACCATAT TTCGGGCCAG CAGCCGAGGG AATTTACATA
      CTGACCGGAC CTATGGTATA AAGCCCGGTC GTCGGCTCCC TTAAATGTAT

3701  GAGGGGCTAA TGCACAATCA AGATGGTTTA ATCTGTGGGT TGAGACAGCT
      CTCCCCGATT ACGTGTTAGT TCTACCAAAT TAGACACCCA ACTCTGTCGA

3751  GGCCAACGAG ACGACTCAAG CTCTTCAACT GTTCCTGAGA GCCACAACTG
      CCGGTTGCTC TGCTGAGTTC GAGAAGTTGA CAAGGACTCT CGGTGTTGAC

3801  AGCTACGCAC CTTTTCAATC CTCAACCGTA AGGCAATTGA TTTCTTGCTG
      TCGATGCGTG GAAAAGTTAG GAGTTGGCAT TCCGTTAACT AAAGAACGAC

3851  CAGCGATGGG GCGGCACATG CCACATTCTG GGACCGGACT GCTGTATCGA
      GTCGCTACCC CGCCGTGTAC GGTGTAAGAC CCTGGCCTGA CGACATAGCT

3901  ACCACATGAT TGGACCAAGA ACATAACAGA CAAAATTGAT CAGATTATTC
      TGGTGTACTA ACCTGGTTCT TGTATTGTCT GTTTTAACTA GTCTAATAAG

3951  ATGATTTTGT TGATAAAACC CTTCCGGACC AGGGGGACAA TGACAATTGG
      TACTAAAACA ACTATTTTGG GAAGGCCTGG TCCCCCTGTT ACTGTTAACC

4001  TGGACAGGAT GGAGACAATG GATACCGGCA GGTATTGGAG TTACAGGCGT
      ACCTGTCCTA CCTCTGTTAC CTATGGCCGT CCATAACCTC AATGTCCGCA

4051  TATAATTGCA GTTATCGCTT TATTCTGTAT ATGCAAATTT GTCTTTTAGT
      ATATTAACGT CAATAGCGAA ATAAGACATA TACGTTTAAA CAGAAAATCA

4101  TTTTCTTCAG ATTGCTTCAT GGAAAAGCTC AGCCTCAAAT CAATGAAACC
      AAAAGAAGTC TAACGAAGTA CCTTTTCGAG TCGGAGTTTA GTTACTTTGG

4151  AGGATTTAAT TATATGGATT ACTTGAATCT AAGATTACTT GACAAATGAT
      TCCTAAATTA ATATACCTAA TGAACTTAGA TTCTAATGAA CTGTTTACTA

4201  AATATAATAC ACTGGAGCTT TAAACATAGC CAATGTGATT CTAACTCCTT
      TTATATTATG TGACCTCGAA ATTTGTATCG GTTACACTAA GATTGAGGAA

4251  TAAACTCACA GTTAATCATA AACAAGGTTT GGAATTGATC TGCTGTGCCT
      ATTTGAGTGT CAATTAGTAT TTGTTCCAAA CCTTAACTAG ACGACACGGA

4301  TCTAGTTGCC AGCCATCTGT TGTTTGCCCC TCCCCCGTGC CTTCCTTGAC
      AGATCAACGG TCGGTAGACA ACAAACGGGG AGGGGGCACG GAAGGAACTG

4351  CCTGGAAGGT GCCACTCCCA CTGTCCTTTC CTAATAAAAT GAGGAAATTG
      GGACCTTCCA CGGTGAGGGT GACAGGAAAG GATTATTTTA CTCCTTTAAC

4401  CATCGCATTG TCTGAGTAGG TGTCATTCTA TTCTGGGGGG TGGGGTGGGG
      GTAGCGTAAC AGACTCATCC ACAGTAAGAT AAGACCCCCC ACCCCACCCC
```

```
                                             SphI
4451  CAGCACAGCA AGGGGGAGGA TTGGGAAGAC AATAGCAGGC ATGCTGGGGA
      GTCGTGTCGT TCCCCCTCCT AACCCTTCTG TTATCGTCCG TACGACCCCT

                 KpnI
4501  TGCGGTGGGC TCTATGGGTA CCCAGGTGCT GAAGAATTGA CCCGGTTCCT
      ACGCCACCCG AGATACCCAT GGGTCCACGA CTTCTTAACT GGGCCAAGGA

4551  CCTGGGCCAG AAAGAAGCAG GCACATCCCC TTCTCTGTGA CACACCCTGT
      GGACCCGGTC TTTCTTCGTC CGTGTAGGGG AAGAGACACT GTGTGGGACA

4601  CCACGCCCCT GGTTCTTAGT TCCAGCCCCA CTCATAGGAC ACTCATAGCT
      GGTGCGGGGA CCAAGAATCA AGGTCGGGGT GAGTATCCTG TGAGTATCGA

4651  CAGGAGGGCT CCGCCTTCAA TCCCACCCGC TAAAGTACTT GGAGCGGTCT
      GTCCTCCCGA GGCGGAAGTT AGGGTGGGCG ATTTCATGAA CCTCGCCAGA

4701  CTCCCTCCCT CATCAGCCCA CCAAACCAAA CCTAGCCTCC AAGAGTGGGA
      GAGGGAGGGA GTAGTCGGGT GGTTTGGTTT GGATCGGAGG TTCTCACCCT

4751  AGAAATTAAA GCAAGATAGG CTATTAAGTG CAGAGGGAGA GAAAATGCCT
      TCTTTAATTT CGTTCTATCC GATAATTCAC GTCTCCCTCT CTTTTACGGA

4801  CCAACATGTG AGGAACTAAT GAGAGAAATC ATAGAATTTC TTCCGCTTCC
      GGTTGTACAC TCCTTGATTA CTCTCTTTAG TATCTTAAAG AAGGCGAAGG

4851  TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC
      AGCGAGTGAC TGAGCGACGC GAGCCAGCAA GCCGACGCCG CTCGCCATAG

4901  AGCTCACTCA AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG
      TCGAGTGAGT TTCCGCCATT ATGCCAATAG GTGTCTTAGT CCCCTATTGC

4951  CAGGAAAGAA CATGTGAGCA AAAGGCCAGC AAAAGGCCAG GAACCGTAAA
      GTCCTTTCTT GTACACTCGT TTTCCGGTCG TTTTCCGGTC CTTGGCATTT

5001  AAGGCCGCGT TGCTGGCGTT TTTCCATAGG CTCCGCCCCC CTGACGAGCA
      TTCCGGCGCA ACGACCGCAA AAAGGTATCC GAGGCGGGGG GACTGCTCGT

5051  TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG ACAGGACTAT
      AGTGTTTTTA GCTGCGAGTT CAGTCTCCAC CGCTTTGGGC TGTCCTGATA

5101  AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT
      TTTCTATGGT CCGCAAAGGG GGACCTTCGA GGGAGCACGC GAGAGGACAA

5151  CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG
      GGCTGGGACG GCGAATGGCC TATGGACAGG CGGAAAGAGG GAAGCCCTTC

5201  CGTGGCGCTT TCTCAATGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG
      GCACCGCGAA AGAGTTACGA GTGCGACATC CATAGAGTCA AGCCACATCC

5251  TCGTTCGCTC CAAGCTGGGC TGTGTGCACG AACCCCCCGT TCAGCCCGAC
      AGCAAGCGAG GTTCGACCCG ACACACGTGC TTGGGGGGCA AGTCGGGCTG

5301  CGCTGCGCCT TATCCGGTAA CTATCGTCTT GAGTCCAACC CGGTAAGACA
      GCGACGCGGA ATAGGCCATT GATAGCAGAA CTCAGGTTGG GCCATTCTGT
```

```
5351  CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT AGCAGAGCGA
      GCTGAATAGC GGTGACCGTC GTCGGTGACC ATTGTCCTAA TCGTCTCGCT

5401  GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC
      CCATACATCC GCCACGATGT CTCAAGAACT TCACCACCGG ATTGATGCCG

5451  TACACTAGAA GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC
      ATGTGATCTT CCTGTCATAA ACCATAGACG CGAGACGACT TCGGTCAATG

5501  CTTCGGAAAA AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG
      GAAGCCTTTT TCTCAACCAT CGAGAACTAG GCCGTTTGTT TGGTGGCGAC

5551  GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC AGATTACGCG CAGAAAAAAA
      CATCGCCACC AAAAAAACAA ACGTTCGTCG TCTAATGCGC GTCTTTTTTT

5601  GGATCTCAAG AAGATCCTTT GATCTTTTCT ACGGGGTCTG ACGCTCAGTG
      CCTAGAGTTC TTCTAGGAAA CTAGAAAAGA TGCCCCAGAC TGCGAGTCAC

5651  GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA TCAAAAAGGA
      CTTGCTTTTG AGTGCAATTC CCTAAAACCA GTACTCTAAT AGTTTTTCCT

5701  TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA
      AGAAGTGGAT CTAGGAAAAT TTAATTTTTA CTTCAAAATT TAGTTAGATT

5751  AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA
      TCATATATAC TCATTTGAAC CAGACTGTCA ATGGTTACGA ATTAGTCACT

5801  GGCACCTATC TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC
      CCGTGGATAG AGTCGCTAGA CAGATAAAGC AAGTAGGTAT CAACGGACTG

5851  TCCGGGGGGG GGGGGCGCTG AGGTCTGCCT CGTGAAGAAG GTGTTGCTGA
      AGGCCCCCCC CCCCCGCGAC TCCAGACGGA GCACTTCTTC CACAACGACT

5901  CTCATACCAG GCCTGAATCG CCCCATCATC CAGCCAGAAA GTGAGGGAGC
      GAGTATGGTC CGGACTTAGC GGGGTAGTAG GTCGGTCTTT CACTCCCTCG

5951  CACGGTTGAT GAGAGCTTTG TTGTAGGTGG ACCAGTTGGT GATTTTGAAC
      GTGCCAACTA CTCTCGAAAC AACATCCACC TGGTCAACCA CTAAAACTTG

6001  TTTTGCTTTG CCACGGAACG GTCTGCGTTG TCGGGAAGAT GCGTGATCTG
      AAAACGAAAC GGTGCCTTGC CAGACGCAAC AGCCCTTCTA CGCACTAGAC

6051  ATCCTTCAAC TCAGCAAAAG TTCGATTTAT TCAACAAAGC CGCCGTCCCG
      TAGGAAGTTG AGTCGTTTTC AAGCTAAATA AGTTGTTTCG GCGGCAGGGC

6101  TCAAGTCAGC GTAATGCTCT GCCAGTGTTA CAACCAATTA ACCAATTCTG
      AGTTCAGTCG CATTACGAGA CGGTCACAAT GTTGGTTAAT TGGTTAAGAC

6151  ATTAGAAAAA CTCATCGAGC ATCAAATGAA ACTGCAATTT ATTCATATCA
      TAATCTTTTT GAGTAGCTCG TAGTTTACTT TGACGTTAAA TAAGTATAGT

6201  GGATTATCAA TACCATATTT TTGAAAAAGC CGTTTCTGTA ATGAAGGAGA
      CCTAATAGTT ATGGTATAAA AACTTTTTCG GCAAAGACAT TACTTCCTCT

6251  AAACTCACCG AGGCAGTTCC ATAGGATGGC AAGATCCTGG TATCGGTCTG
      TTTGAGTGGC TCCGTCAAGG TATCCTACCG TTCTAGGACC ATAGCCAGAC
```

```
6301  CGATTCCGAC TCGTCCAACA TCAATACAAC CTATTAATTT CCCCTCGTCA
      GCTAAGGCTG AGCAGGTTGT AGTTATGTTG GATAATTAAA GGGGAGCAGT

6351  AAAATAAGGT TATCAAGTGA GAAATCACCA TGAGTGACGA CTGAATCCGG
      TTTTATTCCA ATAGTTCACT CTTTAGTGGT ACTCACTGCT GACTTAGGCC

                  HindIII
6401  TGAGAATGGC AAAAGCTTAT GCATTTCTTT CCAGACTTGT TCAACAGGCC
      ACTCTTACCG TTTTCGAATA CGTAAAGAAA GGTCTGAACA AGTTGTCCGG

6451  AGCCATTACG CTCGTCATCA AAATCACTCG CATCAACCAA ACCGTTATTC
      TCGGTAATGC GAGCAGTAGT TTTAGTGAGC GTAGTTGGTT TGGCAATAAG

                                         PvuI
6501  ATTCGTGATT GCGCCTGAGC GAGACGAAAT ACGCGATCGC TGTTAAAAGG
      TAAGCACTAA CGCGGACTCG CTCTGCTTTA TGCGCTAGCG ACAATTTTCC

6551  ACAATTACAA ACAGGAATCG AATGCAACCG GCGCAGGAAC ACTGCCAGCG
      TGTTAATGTT TGTCCTTAGC TTACGTTGGC CGCGTCCTTG TGACGGTCGC

6601  CATCAACAAT ATTTTCACCT GAATCAGGAT ATTCTTCTAA TACCTGGAAT
      GTAGTTGTTA TAAAAGTGGA CTTAGTCCTA TAAGAAGATT ATGGACCTTA

6651  GCTGTTTTCC CGGGGATCGC AGTGGTGAGT AACCATGCAT CATCAGGAGT
      CGACAAAAGG GCCCCTAGCG TCACCACTCA TTGGTACGTA GTAGTCCTCA

6701  ACGGATAAAA TGCTTGATGG TCGGAAGAGG CATAAATTCC GTCAGCCAGT
      TGCCTATTTT ACGAACTACC AGCCTTCTCC GTATTTAAGG CAGTCGGTCA

6751  TTAGTCTGAC CATCTCATCT GTAACATCAT TGGCAACGCT ACCTTTGCCA
      AATCAGACTG GTAGAGTAGA CATTGTAGTA ACCGTTGCGA TGGAAACGGT

6801  TGTTTCAGAA ACAACTCTGG CGCATCGGGC TTCCCATACA ATCGATAGAT
      ACAAAGTCTT TGTTGAGACC GCGTAGCCCG AAGGGTATGT TAGCTATCTA

6851  TGTCGCACCT GATTGCCCGA CATTATCGCG AGCCCATTTA TACCCATATA
      ACAGCGTGGA CTAACGGGCT GTAATAGCGC TCGGGTAAAT ATGGGTATAT

                                        XhoI
6901  AATCAGCATC CATGTTGGAA TTTAATCGCG GCCTCGAGCA AGACGTTTCC
      TTAGTCGTAG GTACAACCTT AAATTAGCGC CGGAGCTCGT TCTGCAAAGG

6951  CGTTGAATAT GGCTCATAAC ACCCCTTGTA TTACTGTTTA TGTAAGCAGA
      GCAACTTATA CCGAGTATTG TGGGGAACAT AATGACAAAT ACATTCGTCT

7001  CAGTTTTATT GTTCATGATG ATATATTTTT ATCTTGTGCA ATGTAACATC
      GTCAAAATAA CAAGTACTAC TATATAAAAA TAGAACACGT TACATTGTAG

                  DraIII
7051  AGAGATTTTG AGACACAACG TGGCTTTCCC CCCCCCCCCA TTATTGAAGC
      TCTCTAAAAC TCTGTGTTGC ACCGAAAGGG GGGGGGGGGT AATAACTTCG

7101  ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG AATGTATTTA
      TAAATAGTCC CAATAACAGA GTACTCGCCT ATGTATAAAC TTACATAAAT
```

```
7151  CAAAAATAAA CAAATAGGGG TTCCGCGCAC ATTTCCCCGA AAAGTGCCAC
      GTTTTTATTT GTTTATCCCC AAGGCGCGTG TAAAGGGGCT TTTCACGGTG
..........................................................................
7201  CTGACGTCTA AGAAACCATT ATTATCATGA CATTAACCTA TAAAAATAGG
      GACTGCAGAT TCTTTGGTAA TAATAGTACT GTAATTGGAT ATTTTTATCC
..........................................................................
7251  CGTATCACGA GGCCCTTTCG TC
      GCATAGTGCT CCGGGAAAGC AG
..........................................................................
```

DECLARATION AND POWER OF ATTORNEY- USA PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I believe I am an original, first and joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled IMMUNIZATION FOR EBOLA VIRUS INFECTION; the specification of which was filed on December 23, 1998, as International Application No. PCT/US98/27364, the present application representing the U.S. national phase thereof.

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above;

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, § 1.56;

I hereby claim the benefit under Title 35, United States Codes § 119(e) of any United States provisional application(s) listed below.

Application No.: 60/068,655 Filing Date: December 23, 1997

POWER OF ATTORNEY: I hereby appoint the registrants of Knobbe, Martens, Olson & Bear, LLP, 620 Newport Center Drive, Sixteenth Floor, Newport Beach, California 92660, Telephone (949) 760-0404, **Customer No. 20,995**.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful, false statements may jeopardize the validity of the application or any patent issued thereon.

1-00 Full name of first inventor: **Gary J. Nabel**

Inventor's signature Gary J Nabel

Date 7/20/01

Residence: **2520 30th St., NW, Washington, D.C. 20008**

DC

Citizenship: **U.S.**

Full name of second inventor: **Anthony Sanchez**

Inventor's signature ____________________

Date July 20, 2001

2.00

Residence: **1717 Red Fox Run, Lilbu~~m~~rn, Atlanta, GA 30047**

GA

Citizenship: **U.S.**

Send Correspondence To:
KNOBBE, MARTENS, OLSON & BEAR, LLP
Customer No. 20,995

FOR-69P
O:\DOCS\NWV\NWV-4776.DOC
070301